WORLD INTELLECTUAL PROPERTY OF International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6:

C12N 15/31, 1/21, C12P 21/02, C07K 14/33, A61K 38/16, 39/08

(11) International Publication Number:

NIZATION

WO 98/07864

A1

(43) International Publication Date: 26 February 1998 (26.02.98)

(21) International Application Number:

PCT/GB97/02273

(22) International Filing Date:

22 August 1997 (22.08.97)

(30) Priority Data:

9617671.4

23 August 1996 (23.08.96) GB

9625996.5

13 December 1996 (13.12.96) GB

(60) Parent Application or Grant

(63) Related by Continuation

US Filed on 08/782,893 (CIP)

27 December 1996 (27.12.96)

(71) Applicants (for all designated States except US): MICRO-BIOLOGICAL RESEARCH AUTHORITY CAMR (CEN-TRE FOR APPLIED MICROBIOLOGY & RESEARCH) [GB/GB]; Porton Down, Salisbury, Wiltshire SP4 0JG (GB). THE SPEYWOOD LABORATORY LIMITED [GB/GB]; 14 Kensington Square, London W8 5HH (GB).

(72) Inventors; and

(75) Inventors/Applicants (for US only): SHONE, Clifford, Charles [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB). QUINN,

Conrad, Padraig [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB). FOSTER, Keith, Alan [GB/GB]; Microbiological Research Authority CAMR (Centre for Applied Microbiology & Research), Porton Down, Salisbury, Wiltshire SP4 0JG (GB).

(74) Agent: SCHLICH, George, William; Mathys & Squire, 100 Gray's Inn Road, London WC1X 8AL (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD,

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: RECOMBINANT TOXIN FRAGMENTS

(57) Abstract

A polypeptide has first and second domains which enable the polypeptide to be translocated into a target cell or which increase the solubility of the polypeptide, or both, and further enable the polypeptide to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis. The polypeptide thus combines useful properties of a clostridial toxin, such as a botulinum or tetanus toxin, without the toxicity associated with the natural molecule. The polypeptide can also contain a third domain that targets it to a specific cell, rendering the polypeptide useful in inhibition of exocytosis in target cells. Fusion proteins comprising the polypeptide, nucleic acids encoding the polypeptide and methods of making the polypeptide are also provided. Controlled activation of the polypeptide is possible and the polypeptide can be incorporated into vaccines and toxin assays.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	1.atvia	SZ	Swaziland
A7.	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazi)	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of Americ
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Vict Nam
CG	Congo	KE	Кепуа	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KÇ	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ.	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	· SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

RECOMBINANT TOXIN FRAGMENTS

This invention relates to recombinant toxin fragments, to DNA encoding these fragments and to their uses such as in a vaccine and for *in vitro* and *in vivo* purposes.

The clostridial neurotoxins are potent inhibitors of calcium-dependent neurotransmitter secretion in neuronal cells. They are currently considered to mediate this activity through a specific endoproteolytic cleavage of at least one of three vesicle or pre-synaptic membrane associated proteins VAMP, syntaxin or SNAP-25 which are central to the vesicle docking and membrane fusion events of neurotransmitter secretion. The neuronal cell targeting of tetanus and botulinum neurotoxins is considered to be a receptor mediated event following which the toxins become internalised and subsequently traffic to the appropriate intracellular compartment where they effect their endopeptidase activity.

The clostridial neurotoxins share a common architecture of a catalytic L-chain (LC, ca 50 kDa) disulphide linked to a receptor binding and translocating H-chain (HC, ca 100 kDa). The HC polypeptide is considered to comprise all or part of two distinct functional domains. The carboxy-terminal half of the HC (ca 50 kDa), termed the $H_{\rm C}$ domain, is involved in the high affinity, neurospecific binding of th neurotoxin to cell surface receptors on the target neuron, whilst the amino-terminal half, termed the $H_{\rm N}$ domain (ca 50 kDa), is considered to mediate the translocation of at least some portion of the neurotoxin across cellular membranes such that the functional activity of the LC is expressed within the target cell. The $H_{\rm N}$ domain also has the property, under conditions of low pH, of forming ion-permeable channels in lipid membranes, this may in some manner relate to its translocation function.

For botulinum neurotoxin type A (BoNT/A) these domains are considered to reside within amino acid residues 872-1296 for the $H_{\rm c}$, amino acid residues 449-871 for the $H_{\rm N}$ and residues 1-448 for th LC. Digestion with trypsin effectively degrades the $H_{\rm c}$ domain of the BoNT/A to generat a non-toxic fragment designated LH_N,

PCT/GB97/02273

which is no longer able to bind to and enter neurons (Fig. 1). The LH_N fragment so produced also has the property of enhanced solubility compared to both the parent holotoxin and the isolated LC.

It is therefore possible to provide functional definitions of the domains within the neurotoxin molecule, as follows:

- (A) clostridial neurotoxin light chain:
- -a metalloprotease exhibiting high substrate specificity for vesicle and/or plasma membrane associated proteins involved in the exocytotic process. In particular, it cleaves one or more of SNAP-25, VAMP (synaptobrevin / cellubrevin) and syntaxin.
- (B) clostridial neurotoxin heavy chain H_N domain:
- -a portion of the heavy chain which enables translocation of that portion of the neurotoxin molecule such that a functional expression of light chain activity occurs within a target cell.
- -the domain responsible for translocation of the endopeptidase activity, following binding of neurotoxin to its specific cell surface receptor via the binding domain, into the target cell.
- -the domain responsible for formation of ion-permeable pores in lipid membranes under conditions of low pH.
- -the domain responsible for increasing the solubility of the entire polypeptide compared to the solubility of light chain alone.
- (C) clostridial neurotoxin heavy chain H_c domain.

holotoxin to cell surface receptor(s) involved in the intoxicating action of clostridial toxin prior to int rnalisation of the toxin into the cell.

The identity of the cellular recognition markers for these toxins is currently not understood and no specific receptor species have yet been identified although Kozaki et al. have reported that synaptotagmin may be the receptor for botulinum neurotoxin type B. It is probable that each of the neurotoxins has a different receptor.

It is desirable to have positive controls for toxin assays, to develop clostridial toxin vaccines and to develop therapeutic agents incorporating desirable properties of clostridial toxin.

However, due to its extreme toxicity, the handling of native toxin is hazardous.

The present invention seeks to overcome or at least ameliorate problems associated with production and handling of clostridial toxin.

Accordingly, the invention provides a polypeptide comprising first and second domains, wherein said first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to neuronal exocytosis and wherein said second domain is adapted (i) to translocate the polypeptide into the cell or (ii) to increase the solubility of the polypeptide compared to the solubility of the first domain on its own or (iii) both to translocate the polypeptide into the cell and to increase the solubility of the polypeptide compared to the solubility of the first domain on its own, said polypeptide being free of clostridial neurotoxin and fre of any clostridial neurotoxin precursor that can be converted into toxin by proteolytic action. Accordingly, the invention may thus provide a single polypeptide chain containing a domain equivalent to a clostridial toxin light chain and a domain providing the functional aspects of the H_N of a clostridial toxin heavy chain, whilst lacking the functional aspects of a clostridial toxin H_C domain.

- 4 -

For the purposes of the invention, the functional property or properties of the H_N of a clostridial toxin heavy chain that are required to be exhibited by the second domain of the polypeptide of the invention are either (i) translocation of the polypeptide into a cell, or (ii) increasing solubility of the polypeptide compared to solubility of the first domain on its own or (iii) both (i) and (ii). References hereafter

or properties. The second domain is not required to exhibit other properties of the H_a domain of a clostridial toxin heavy chain.

to a H_N domain or to the functions of a H_N domain are references to this property

A polypeptide of the invention can thus be soluble but lack the translocation function of a native toxin-this is of use in providing an immunogen for vaccinating or assisting to vaccinate an individual against challenge by toxin. In a specific embodiment of the invention described in an example below a polypeptide designated LH₄₂₃/A elicited neutralising antibodies against type A neurotoxin. A polypeptide of the invention can likewise thus be relatively insoluble but retain the translocation function of a native toxin - this is of use if solubility is imparted to a composition made up of that polypeptide and one or more other components by one or more of said other components.

The first domain of the polypeptide of the invention cleaves one or more vesicle or plasma-membrane associated proteins essential to the specific cellular process of exocytosis, and cleavage of these proteins results in inhibition of exocytosis, typically in a non-cytotoxic manner. The cell or cells affected are not restricted to a particular type or subgroup but can include both neuronal and non-neuronal cells. The activity of clostridial neurotoxins in inhibiting exocytosis has, indeed, been observed almost universally in eukaryotic cells expressing a relevant cell surface receptor, including such diverse cells as from Aplysia (sea slug), Drosophila (fruit fly) and mammalian nerve cells, and the activity of the first domain is to be understood as including a corresponding range of cells.

The polypeptide of the invention may be obtained by expression of a recombinant nucleic acid, preferably a DNA, and is a single-polypeptide, that is to say not

cleaved into separate light and heavy chain domains. The polypeptide is thus available in convenient and large quantities using recombinant techniques.

In a polypeptide according to the invention, said first domain preferably comprises a clostridial toxin light chain or a fragment or variant of a clostridial toxin light chain. The fragment is optionally an N-terminal, or C-terminal fragment of the light chain, or is an internal fragment, so long as it substantially retains the ability to cleave the vesicle or plasma-membrane associated protein essential to exocytosis. The minimal domains necessary for the activity of the light chain of clostridial toxins are described in J. Biol. Chem., Vol.267, No. 21, July 1992, pages 14721-14729. The variant has a different peptide sequence from the light chain or from the fragment, though it too is capable of cleaving the vesicle or plasma-membrane associated protein. It is conveniently obtained by insertion, deletion and/or substitution of a light chain or fragment thereof. In embodiments of the invention described below a variant sequence comprises (i) an N-terminal extension to a clostridial toxin light chain or fragment (ii) a clostridial toxin light chain or fragment modified by alteration of at least one amino acid (iii) a C-terminal extension to a clostridial toxin light chain or fragment, or (iv) combinations of 2 or more of (i)-(iii).

In further embodiments of the invention, the variant contains an amino acid sequence modified so that (a) there is no protease sensitive region between th LC and H_N components of the polypeptide, or (b) the protease sensitive region is specific for a particular protease. This latter embodiment is of use if it is desired to activate the endopeptidase activity of the light chain in a particular environment or cell. Though, in general, the polypeptides of the invention are activated prior to administration.

The first domain preferably exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin. The clostridial toxin is preferably botulinum toxin or tetanus toxin.

In an embodiment of the invention described in an example below, the toxin light



chain and the portion of the toxin heavy chain are of botulinum toxin type A. In a further embodiment of the invention described in an example below, the toxin light chain and the portion of the toxin heavy chain are of botulinum toxin type B. The polypeptide optionally comprises a light chain or fragment or variant of one toxin type and a heavy chain or fragment or variant of another toxin type.

In a polypeptide according to the invention said second domain preferably comprises a clostridial toxin heavy chain H_N portion or a fragment or variant of a clostridial toxin heavy chain H_N portion. The fragment is optionally an N-terminal or C-terminal or internal fragment, so long as it retains the function of the H_N domain. Teachings of regions within the H_N responsible for its function are provided for example in Biochemistry 1995, 34, pages 15175-15181 and Eur. J. Biochem, 1989, 185, pages 197-203. The variant has a different sequence from the H_N domain or fragment, though it too retains the function of the H_N domain. It is conveniently obtained by insertion, deletion and/or substitution of a H_N domain or fragment thereof. In embodiments of the invention, described below, it comprises (i) an N-terminal extension to a H_N domain or fragment, (iii) a C-terminal extension to a H_N domain or fragment, (iii) a modification to a H_N domain or fragment by alteration of at least one amino acid, or (iv) combinations of 2 or more of (i)-(iii). The clostridial toxin is preferably botulinum toxin or tetanus toxin.

The invention also provides a polypeptide comprising a clostridial neurotoxin light chain and a N-terminal fragment of a clostridial neurotoxin heavy chain, the fragment preferably comprising at least 423 of the N-terminal amino acids of the heavy chain of botulinum toxin type A, 417 of the N-terminal amino acids of the heavy chain of botulinum toxin type B or the equivalent number of N-terminal amino acids of the heavy chain of other types of clostridial toxin such that the fragment possesses an equivalent alignment of homologous amino acid residues.

These polypeptides of the invention are thus not composed of two or more polypeptides, linked for example by di-sulphide bridges into composite molecules.

—Instead, these-polypeptides_are_single_chains and are not active or their activity is

significantly reduced in an in vitro assay of neurotoxin endopeptidase activity.

Further, the polypeptides may be susceptible to be converted into a form exhibiting endopeptidase activity by the action of a proteolytic agent, such as trypsin. In this way it is possible to control the endopeptidase activity of the toxin light chain.

In a specific embodiment of the invention described in an example below, there is provided a polypeptide lacking a portion designated $H_{\rm C}$ of a clostridial toxin heavy chain. This portion, seen in the naturally produced toxin, is responsible for binding of toxin to cell surface receptors prior to internalisation of the toxin. This specific embodiment is therefore adapted so that it can not be converted into active toxin, for example by the action of a proteolytic enzyme. The invention thus also provides a polypeptide comprising a clostridial toxin light chain and a fragment of a clostridial toxin heavy chain, said fragment being not capable of binding to those cell surface receptors involved in the intoxicating action of clostridial toxin, and it is preferred that such a polypeptide lacks an intact portion designated $H_{\rm C}$ of a clostridial toxin heavy chain.

In further embodiments of the invention there are provided compositions containing a polypeptide comprising a clostridial toxin light chain and a portion designated H_N of a clostridial toxin heavy chain, and wherein the composition is free of clostridial toxin and free of any clostridial toxin precursor that may be converted into clostridial toxin by the action of a proteolytic enzyme. Examples of these compositions include those containing toxin light chain and H_N sequences of botulinum toxin types A, B, C₁, D, E, F and G.

The polypeptides of the invention are conveniently adapted to bind to, or include, a ligand for targeting to desired cells. The polypeptide optionally comprises a sequence that binds to, for example, an immunoglobulin. A suitable sequence is a tandem repeat synthetic IgG binding domain derived from domain B of Staphylococcal protein A. Choice of immunoglobulin specificity then determines the target for a polypeptide - immunoglobulin complex. Alternatively, the

polypeptide comprises a non-clostridial sequence that binds to a cell surface receptor, suitable sequences including insulin-like growth factor-1 (IGF-1) which binds to its specific receptor on particular cell types and the 14 amino acid residue sequence from the carboxy-terminus of cholera toxin A subunit which is able to bind the cholera toxin B subunit and thence to GM1 gangliosides. A polypeptide according to the invention thus, optionally, further comprises a third domain adapted for binding of the polypeptide to a cell.

In a second aspect the invention provides a fusion protein comprising a fusion of (a) a polypeptide of the invention as described above with (b) a second polypeptide adapted for binding to a chromatography matrix so as to enable purification of the fusion protein using said chromatography matrix. It is convenient for the second polypeptide to be adapted to bind to an affinity matrix, such as a glutathione Sepharose, enabling rapid separation and purification of the fusion protein from an impure source, such as a cell extract or supernatant.

One possible second purification polypeptide is glutathione-S-transferase (GST), and others will be apparent to a person of skill in the art, being chosen so as to enable purification on a chromatography column according to conventional techniques.

As noted above, by proteolytic treatment, for example using trypsin, of a polypeptide of the invention it is possible to induce endopeptidase activity in the treated polypeptide. A third aspect of the invention provides a composition comprising a derivative of a clostridial toxin, said derivative retaining at least 10% of the endopeptidase activity of the clostridial toxin, said derivative further being non-toxic *in vivo* due to its inability to bind to cell surface receptors, and wherein the composition is free of any component, such as toxin or a further toxin derivative, that is toxic *in vivo*. The activity of the derivative preferably approaches that of natural toxin, and is thus preferably at least 30% and most preferably at least 60% of natural toxin. The overall indopeptidase activity of the composition will, of course, also be differentiable to induce the derivative that is present.

١

٠,

While it is known to treat naturally produced clostridial toxin to remove the H_C domain, this treatment does not totally remove toxicity of the preparation, instead some residual toxin activity remains. Natural toxin treated in this way is therefore still not entirely safe. The composition of the invention, derived by treatment of a pure source of polypeptide advantageously is free of toxicity, and can conveniently be used as a positive control in a toxin assay, as a vaccine against clostridial toxin or for other purposes where it is essential that there is no residual toxicity in the composition.

The invention enables production of the polypeptides and fusion proteins of the invention by recombinant means.

A fourth aspect of the invention provides a nucleic acid encoding a polypeptide or a fusion protein according to any of the aspects of the invention described above.

In one embodiment of this aspect of the invention, a DNA sequence provided to code for the polypeptide or fusion protein is not derived from native clostridial sequences, but is an artificially derived sequence not preexisting in nature.

A specific DNA (SEQ ID NO: 1) described in more detail below encodes a polypeptide or a fusion protein comprising nucleotides encoding residues 1-871 of a botulinum toxin type A. Said polypeptide comprises the light chain domain and the first 423 amino acid residues of the amino terminal portion of a botulinum toxin type A heavy chain. This recombinant product is designated LH₄₂₃/A (SEQ ID NO: 2).

In a second embodiment of this aspect of the invention a DNA sequence which codes for the polypeptide or fusion protein is derived from native clostridial sequences but codes for a polypeptide or fusion protein not found in nature.

A specific DNA (SEQ ID NO: 19) described in more detail below encodes a polypeptide or a fusion protein and comprises nucleotides encoding residues 1-



1171 of a botulinum toxin type B. Said polypeptide comprises the light chain domain and the first 728 amino acid residues of the amino terminal protein of a botulinum type B heavy chain. This recombinant product is designated LH_{728}/B (SEQ ID NO: 20).

- 10 -

The invention thus also provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA according to the third aspect of the invention. The host cell is suitably not able to cleave a polypeptide or fusion protein of the invention so as to separate light and heavy toxin chains; for example, a non-clostridial host.

The invention further provides a method of manufacture of a polypeptide comprising expressing in a host cell a DNA encoding a fusion protein as described above, purifying the fusion protein by elution through a chromatography column adapted to retain the fusion protein, eluting through said chromatography column a ligand adapted to displace the fusion protein and recovering the fusion protein. Production of substantially pure fusion protein is thus made possible. Likewise, th fusion protein is readily cleaved to yield a polypeptide of the invention, again in substantially pure form, as the second polypeptide may conveniently be removed using the same type of chromatography column.

The LH_N/A derived from dichain native toxin requires extended digestion with trypsin to remove the C-terminal 1/2 of the heavy chain, the H_C domain. The loss of this domain effectively renders the toxin inactive *in vivo* by preventing its interaction with host target cells. There is, however, a residual toxic activity which may indicate a contaminating, trypsin insensitive, form of the whole type A neurotoxin.

In contrast, the recombinant preparations of the invention are the product of a discreet, defined gene coding sequence and can not be contaminated by full length toxin protein. Furthermore, the product as recovered from *E. coli*, and from other recombinant-expression-hosts, is an inactive single chain peptide or if expression-

hosts produce a proc ssed, active polypeptide it is not a toxin. Endopeptidase activity of LH₄₂₃/A, as assess d by the current *in vitro* peptide cleavage assay, is wholly dependent on activation of the recombinant molecule between residues 430 and 454 by trypsin. Other proteolytic enzymes that cleave between these two residues are generally also suitable for activation of the recombinant molecule. Trypsin cleaves the peptide bond C-terminal to Arginine or C-terminal to Lysine and is suitable as these residues are found in the 430-454 region and are exposed (see Fig. 12).

The recombinant polypeptides of the invention are potential therapeutic agents for targeting to cells expressing the relevant substrate but which are not implicated in effecting botulism. An example might be where secretion of neurotransmitter is inappropriate or undesirable or alternatively where a neuronal cell is hyperactive in terms of regulated secretion of substances other than neurotransmitter. In such an example the function of the H_C domain of the native toxin could be replaced by an alternative targeting sequence providing, for example, a cell receptor ligand and/or translocation domain.

One application of the recombinant polypeptides of the invention will be as a reagent component for synthesis of therapeutic molecules, such as disclosed in WO-A-94/21300. The recombinant product will also find application as a non-toxic standard for the assessment and development of *in vitro* assays for detection of functional botulinum or tetanus neurotoxins either in foodstuffs or in environmental samples, for example as disclosed in EP-A-0763131.

A further option is addition, to the C-terminal end of a polypeptide of the invention, of a peptide sequence which allows specific chemical conjugation to targeting ligands of both protein and non-protein origin.

In yet a further embodiment an alternative targeting ligand is added to the N-terminus of polypeptides of the invention. Recombinant LH_N derivatives have been designated that have specific protease cleavage sites engineered at the C-terminus

of the LC at the putative trypsin s nsitive region and also at the extreme C-terminus of the complete protein product. These sites will enhance the activational specificity of the recombinant product such that the dichain species can only be activated by proteolytic cleavage of a more predictable nature than use of trypsin.

The LH_N enzymatically produced from native BoNT/A is an efficient immunogen and thus the recombinant form with its total divorce from any full length neurotoxin represents a vaccine component. The recombinant product may serve as a basal reagent for creating defined protein modifications in support of any of the above areas.

Recombinant constructs are assigned distinguishing names on the basis of their amino acid sequence length and their Light Chain (L-chain, L) and Heavy Chain (H-chain, H) content as these relate to translated DNA sequences in the public domain or specifically to SEQ ID NO: 2 and SEQ ID NO: 20. The 'LH' designation is followed by '/X' where 'X' denotes the corresponding clostridial toxin serotype or class, e.g. 'A' for botulinum neurotoxin type A or 'TeTx' for tetanus toxin. Sequence variants from that of the native toxin polypeptide are given in parenthesis in standard format, namely the residue position number prefixed by the residue of the native sequence and suffixed by the residue of the variant.

Subscript number prefixes indicate an amino-terminal (N-terminal) extension, or where negative a deletion, to the translated sequence. Similarly, subscript number suffixes indicate a carboxy terminal (C-terminal) extension or where negative numbers are used, a deletion. Specific sequence inserts such as protease cleavage sites are indicated using abbreviations, e.g. Factor Xa is abbreviated to FXa. L-chain C-terminal suffixes and H-chain N-terminal prefixes are separated by a / to indicate the predicted junction between the L and H-chains. Abbreviations for engineered ligand sequences are prefixed or suffixed to the clostridial L-chain or H-chain corresponding to their position in the translation product.

Following this_nomenclature,_

SEQ ID NO: 2, containing the entire L-chain and 423 LH₄₂₃/A amino acids of the H-chain of botulinum neurotoxin type A; a variant of this molecule, containing a two amino acid 2LH423/A extension to the N-terminus of the L-chain: a further variant in which the molecule contains a two 2L12H423/A amino acid extension on the N-terminus of both the Lchain and the H-chain: 2LFXa/2H423/A a further variant containing a two amino acid extension to the N-terminus of the L-chain, and a Factor Xa cleavage sequence at the C-terminus of the L-chain which, after cleavage of the molecule with Factor Xa leaves a two amino acid N-terminal extension to the Hchain component; and

 $_2L_{FXa/2}H_{423}/A$ -IGF-1 = a variant of this molecule which has a further C-terminal extension to the H-chain, in this example the insulin-like growth factor 1 (IGF-1) sequence.

There now follows description of specific embodiments of the invention, illustrated by drawings in which:

Fig. 1 shows a schematic representation of the domain structure of botulinum neurotoxin type A (BoNT/A);

Fig. 2 shows a schematic representation of assembly of the gene for an embodiment of the invention designated LH_{423}/A ;

- Fig. 3 is a graph comparing activity of native toxin, trypsin generated "native" LH_N/A and an embodiment of the invention designat d $_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) in an *in vitro* peptide cleavage assay;
- Fig. 4 is a comparison of the first 33 amino acids in published sequences of native toxin and embodiments of the invention;
- Fig. 5 shows the transition region of an embodiment of the invention designated L/₄H₄₂₃/A illustrating insertion of four amino acids at the N-terminus of the H_N sequence; amino acids coded for by the *Eco* 47 III restriction endonuclease cleavage site are marked and the H_N sequence then begins ALN...;
- Fig. 6 shows the transition region of an embodiment of the invention designated $L_{FXa/3}H_{423}/A$ illustrating insertion of a Factor Xa cleavage site at the C-terminus of the L-chain, and three additional amino acids coded for at the N-terminus of the H-sequence; the N-terminal amino acid of the cleavage-activated H_N will be cysteine;
- Fig. 7 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated $L_{FXa/3}H_{423}/A$ -IGF-1, a fusion protein; the IGF-1 sequence begins at position G_{882} ;
- Fig. 8 shows the C-terminal portion of the amino acid sequence of an embodiment of the invention designated $L_{FXa/3}H_{423}/A$ -CtxA14, a fusion protein; the C-terminal CtxA sequence begins at position Ω_{882} ;
- Fig.9 ——shows-the-C-terminal-portion of-the-amino-acid-sequence-ofan-

. 4

embodiment of the invention designated $L_{FXa/3}H_{423}/A-ZZ$, a fusion protein; th C-terminal ZZ sequence begins at position A_{890} immediately after a generase recognition site (underlined);

show schematic representations of manipulations of

Figs. 10 & 11

polypeptides of the invention; Fig. 10 shows LH₄₂₃/A with N-terminal addition of an affinity purification peptide (in this case GST) and C-terminal addition of an Ig binding domain; protease cleavage sites R1, R2 and R3 enable selective enzymatic separation of domains; Fig. 11 shows specific examples of protease cleavage sites R1, R2 and R3 and a C-terminal fusion peptide sequence;

Fig. 12

shows the trypsin sensitive activation region of a polypeptide of the invention;

Fig. 13

shows Western blot analysis of recombinant LH₁₀₇/B expressed from *E.coli*; panel A was probed with anti-BoNT/B antiserum; Lane 1, molecular weight standards; lanes 2 & 3, native BoNT/B; lane 4, immunopurified LH₁₀₇/B; panel B was probed with anti-T7 peptide tag antiserum; lane 1, molecular weight standards; lanes 2 & 3, positive control *E.coli* T7 expression; lane 4 immunopurified LH₁₀₇/B.

The sequence listing that accompanies this application contains the following sequences:-

SEQ ID NO: Sequence

1 DNA coding for LH₄₂₃/A

PCT/GB97/02273

2	LH ₄₂₃ /A
3	DNA coding for $_{23}LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$), of which an N-terminal portion is shown in Fig. 4.
4	₂₃ LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y)
5	DNA coding for $_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$), of which an N-
	terminal portion is shown in Fig.4
6	₂ LH ₄₂₃ /A (Q ₂ E ₁ N ₂₆ K ₁ A ₂₇ Y)
	2=423
7 .	DNA coding for native BoNT/A according to Binz et al
8	native BoNT/A according to Binz et al
9	DNA coding for L ₁₄ H ₄₂₃ /A
10	L _{/4} H ₄₂₃ /A
11	DNA coding for L _{FXa} / ₃ H ₄₂₃ /A
12	L _{FXa} / ₃ H ₄₂₃ /A
13	DNA coding for L _{F,Xa} / ₃ H ₄₂₃ /A-IGF-1
14	L _{FXa} / ₃ H ₄₂₃ /A-IGF-1
15	DNA coding for L _{FXa} / ₃ H ₄₂₃ /A-CtxA14
16	L _{FXa} / ₃ H ₄₂₃ /A-CtxA14
17	DNA coding for L _{FXa/3} H ₄₂₃ /A-ZZ
18	$L_{FXa/3}H_{423}/A-ZZ$
19	DNA coding for LH ₇₂₈ /B
20	LH ₇₂₈ /B
21	DNA coding for LH ₄₁₇ /B
22	LH ₄₁₇ /B
23	DNA coding for LH ₁₀₇ /B
24	LH ₁₀₇ /B
25	DNA coding for LH ₄₂₃ /A (Q ₂ E,N ₂₆ K,A ₂₇ Y)
26	$LH_{423}/A (Q_2E, N_{26}K, A_{27}Y)$
 27-	-DNA-coding-for-LH ₄₁₇ /B-wherein-the-first-274-bases-are

WO 98/07864 PCT/GB97/02273

- 17 -

28

. ;

modified to have an *E. coli* codon bias

DNA coding for LH₄₁₇/B wherein bases 691-1641 of the native BoNT/B sequence have been replaced by a degenerate DNA coding for amino acid residues 231-547 of the native BoNT/B polypeptide

Example 1

A 2616 base pair, double stranded gene sequence (SEQ ID NO: 1) has been assembled from a combination of synthetic, chromosomal and polymerase-chain-reaction generated DNA (Figure 2). The gene codes for a polypeptide of 871 amino acid residues corresponding to the entire light-chain (LC, 448 amino acids) and 423 residues of the amino terminus of the heavy-chain (H_c) of botulinum neurotoxin type A. This recombinant product is designated the LH₄₂₃/A fragment (SEQ ID NO: 2).

Construction of the recombinant product

The first 918 base pairs of the recombinant gene were synthesised by concatenation of short oligonucleotides to generate a coding sequence with an E. coli codon bias. Both DNA strands in this region were completely synthesised as short overlapping oligonucleotides which were phosphorylated, annealed and ligated to generate the full synthetic region ending with a unique Kpnl restriction site. The remainder of the LH_{423}/A coding sequence was PCR amplified from total chromosomal DNA from $Clostridium\ botulinum\$ and annealed to the synthetic portion of the gene.

The internal PCR amplified product sequences were then deleted and replaced with the native, fully sequenced, regions from clones of *C. botulinum* chromosomal origin to generate the final gene construct. The final composition is synthetic DNA (bases 1-913), polymerase amplified DNA (bases 914-1138 and 1976-2616) and the remainder is of *C. botulinum* chromosomal origin (bases 1139-1975). The

assembled gene was then fully sequenced and cloned into a variety of *E.coli* plasmid vectors for expression analysis.

Expression of the recombinant gene and recovery of protein product

The DNA is expressed in *E. coli* as a single nucleic acid transcript producing a soluble single chain polypeptide of 99,951 Daltons predicted molecular weight. The gene is currently expressed in *E. coli* as a fusion to the commercially available coding sequence of glutathione S-transferase (GST) of *Schistosoma japonicum* but any of an extensive range of recombinant gene expression vectors such as pEZZ18, pTrc99, pFLAG or the pMAL series may be equally effective as might expression in other prokaryotic or eukaryotic hosts such as the Gram positive bacilli, the yeast *P. pastoris* or in insect or mammalian cells under appropriate conditions.

Currently, E. coli harbouring the expression construct is grown in Luria-Bertani broth (L-broth pH 7.0, containing 10 g/l bacto-tryptone, 5 g/l bacto-yeast extract and 10 g/l sodium chloride) at 37° C until the cell density (biomass) has an optical absorbance of 0.4- 0.6 at 600 nm and the cells are in mid-logarithmic growth phase. Expression of the gene is then induced by addition isopropylthio-β-D-galactosidase (IPTG) to a final concentration of 0.5 mM. Recombinant gene expression is allowed to proceed for 90 min at a reduced The cells are then harvested by centrifugation, are temperature of 25°C. resuspended in a buffer solution containing 10 mM Na₂HPO₄, 0.5 M NaCl, 10 mM EGTA, 0.25% Tween, pH 7.0 and then frozen at -20°C. For extraction of the recombinant protein the cells are_disrupted by sonication. The cell extract is then cleared of debris by centrifugation and the cleared supernatant fluid containing soluble recombinant fusion protein (GST- LH₄₂₃/A) is stored at -20°C pending purification. A proportion of recombinant material is not released by the sonication procedure and this probably reflects insolubility or inclusion body formation. Currently we do not extract this material for analysis but if desired this could be readily achieved using methods known to those skilled in the art.

The recombinant GST- LH_{423}/A is purified by adsorption onto a commercially prepar d affinity matrix of glutathione Sepharose and subsequent elution with reduced glutathione. The GST affinity purification marker is then removed by proteolytic cleavage and reabsorption to glutathione Sepharose; recombinant LH_{423}/A is recovered in the non-adsorbed material.

Construct variants

A variant of the molecule, LH_{423}/A ($Q_2E,N_{26}K,A_{27}Y$) (SEQ ID NO: 26) has been produced in which three amino acid residues have been modified within the light chain of LH_{423}/A producing a polypeptide containing a light chain sequence different to that of the published amino acid sequence of the light chain of BoNT/A.

Two further variants of the gene sequence that have been expressed and the corresponding products purified are $_{23}LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) (SEQ ID NO: 4) which has a 23 amino acid N-terminal extension as compared to the predicted native L-chain of BoNT/A and $_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) (SEQ ID NO: 6) which has a 2 amino acid N-terminal extension (Figure 4).

In yet another variant a gene has been produced which contains a Eco 47 III restriction site between nucleotides 1344 and 1345 of the gene sequence given in (SEQ ID NO: 1). This modification provides a restriction site at the position in the gene representing the interface of the heavy and light chains in native neurotoxin, and provides the capability to make insertions at this point using standard restriction enzyme methodologies known to those skilled in the art. It will also be obvious to those skilled in the art that any one of a number of restriction sites could be so employed, and that the Eco 47 III insertion simply exemplifies this approach. Similarly, it would be obvious for one skilled in the art that insertion of a restriction site in the manner described could be performed on any gene of the invention. The gene described, when expressed, codes for a polypeptide, $L_{/4}H_{423}/A$ (SEQ ID NO: 10), which contains an additional four amino acids between amino acids 448 and 449 of LH_{423}/A at a position equivalent to the amino-terminus-of-the-

heavy chain of native BoNT/A.

A variant of the gene has been expressed, L_{FXa/3}H₄₂₃/A (SEQ ID NO: 12), in which a specific proteolytic cleavage site was incorporated at the carboxy-terminal end of the light chain domain, specifically after residue 448 of L_{/4}H₄₂₃/A. The cleavage site incorporated was for Factor Xa protease and was coded for by modification of SEQ ID NO: 1. It will be apparent to one skilled in the art that a cleavage site for another specified protease could be similarly incorporated, and that any gene sequence coding for the required cleavage site could be employed. Modification of the gene sequence in this manner to code for a defined protease site could be performed on any gene of the invention.

Variants of $L_{FXa/3}H_{423}/A$ have been constructed in which a third domain is present at the carboxy-terminal end of the polypeptide which incorporates a specific binding activity into the polypeptide.

Specific examples described are:

- (1) $L_{FXa/3}H_{423}/A$ -IGF-1 (SEQ ID NO: 14), in which the carboxy-terminal domain has a sequence equivalent to that of insulin-like growth factor-1 (IGF-1) and is able to bind to the insulin-like growth factor receptor with high affinity;
- (2) $L_{FXa/3}H_{423}/A$ -CtxA14 (SEQ ID NO: 16), in which the carboxy-terminal domain has a sequence equivalent to that of the 14 amino acids from the carboxy-terminus of the A-subunit of cholera toxin (CtxA) and is thereby able to interact with the cholera toxin B-subunit pentamer; and
- (3) L_{FXa/3}H₄₂₃/A-ZZ (SEQ ID NO: 18), in which the carboxy-terminal domain is a tandem repeating synthetic IgG binding domain. This variant also exemplifies another modification applicable to the current invention, namely the inclusion in the gene of a sequence coding for a protease cl avage site located between the end of-the-clostridial-heavy-chain-sequence-and-the-sequence-coding-for-the-binding-

- 21 -

ligand. Specifically in this example a sequence is inserted at nucleotides 2650 to 2666 coding for a genenase cleavage site. Expression of this gene produces a polypeptide which has the desired protease sensitivity at the interface between the domain providing H_N function and the binding domain. Such a modification enables selective removal of the C-terminal binding domain by treatment of the polypeptide with the relevant protease.

It will be apparent that any one of a number of such binding domains could be incorporated into the polypeptide sequences of this invention and that the above examples are merely to exemplify the concept. Similarly, such binding domains can be incorporated into any of the polypeptide sequences that are the basis of this invention. Further, it should be noted that such binding domains could be incorporated at any appropriate location within the polypeptide molecules of the invention.

Further embodiments of the invention are thus illustrated by a DNA of the invention further comprising a desired restriction endonuclease site at a desired location and by a polypeptide of the invention further comprising a desired protease cleavage site at a desired location.

The restriction endonuclease site may be introduced so as to facilitate further manipulation of the DNA in manufacture of an expression vector for expressing a polypeptide of the invention; it may be introduced as a consequence of a previous step in manufacture of the DNA; it may be introduced by way of modification by insertion, substitution or deletion of a known sequence. The consequence of modification of the DNA may be that the amino acid sequence is unchanged, or may be that the amino acid sequence is changed, for example resulting in introduction of a desired protease cleavage site, either way the polypeptide retains its first and second domains having the properties required by the invention.

Figure 10 is a diagrammatic representation of an expression product exemplifying features described in this example. Specifically, it illustrates a single-polypeptide-

incorporating a domain equivalent to the light chain of botulinum neurotoxin type A and a domain equivalent to the H_N domain of the h avy chain of botulinum neurotoxin type A with a N-terminal extension providing an affinity purification domain, namely GST, and a C-terminal extension providing a ligand binding domain, namely an IgG binding domain. The domains of the polypeptide are spatially separated by specific protease cleavage sites enabling selective enzymatic separation of domains as exemplified in the Figure. This concept is more specifically depicted in Figure 11 where the various protease sensitivities are defined for the purpose of example.

Assay of product activity

The LC of botulinum neurotoxin type A exerts a zinc-dependent endopeptidase activity on the synaptic vesicle associated protein SNAP-25 which it cleaves in a specific manner at a single peptide bond. The $_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) (SEQ ID NO: 6) cleaves a synthetic SNAP-25 substrate *in vitro* under the same conditions as the native toxin (Figure 3). Thus, the modification of the polypeptide sequence of $_2LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) relative to the native sequence and within the minimal functional LC domains does not prevent the functional activity of the LC domains.

This activity is dependent on proteolytic modification of the recombinant $GST_{-2}LH_{423}/A$ ($Q_2E,N_{26}K,A_{27}Y$) to convert the single chain polypeptide product to a disulphide linked dichain species. This is currently done using the proteolytic enzyme trypsin. The recombinant product (100-600 μ g/ml) is incubated at 37°C for 10-50 minutes with trypsin (10 μ g/ml) in a solution containing 140 mM NaCl, 2.7 mM KCl, 10 mM Na $_2$ HPO $_4$, 1.8 mM KH $_2$ PO $_4$, pH 7.3. The reaction is terminated by addition of a 100-fold molar excess of trypsin inhibitor. The activation by trypsin generates a disulphide linked dichain species as determined by polyacrylamide gel electrophoresis and immunoblotting analysis using polyclonal anti-botulinum neurotoxin type A antiserum.

2LH₄₂₃/A is mor_stabl_in_the_pres_nce_of_trypsin_and_more_active_in_the_in_vitro_

٠٠.

peptide cleavage assay than is ₂₃LH₄₂₃/A. Both variants, however, are fully functional in the *in vitro* peptide cleavage assay. This demonstrates that the r combinant molecule will tolerate N-terminal amino acid extensions and this may be expanded to other chemical or organic moieties as would be obvious to those skilled in the art.

Example 2

As a further exemplification of this invention a number of gene sequences have been assembled coding for polypeptides corresponding to the entire light-chain and varying numbers of residues from the amino terminal end of the heavy chain of botulinum neurotoxin type B. In this exemplification of the disclosure the gene sequences assembled were obtained from a combination of chromosomal and polymerase-chain-reaction generated DNA, and therefore have the nucleotide sequence of the equivalent regions of the natural genes, thus exemplifying the principle that the substance of this disclosure can be based upon natural as well as a synthetic gene sequences.

The gene sequences relating to this example were all assembled and expressed using methodologies as detailed in Sambrook J, Fritsch E F & Maniatis T (1989) Molecular Cloning: A Laboratory Manual (2nd Edition), Ford N, Nolan C, Ferguson M & Ockler M (eds), Cold Spring Harbor Laboratory Press, New York, and known to those skilled in the art.

A gene has been assembled coding for a polypeptide of 1171 amino acids corresponding to the entire light-chain (443 amino acids) and 728 residues from the amino terminus of the heavy chain of neurotoxin type B. Expression of this gene produces a polypeptide, LH_{728}/B (SEQ ID NO: 20), which lacks the specific neuronal binding activity of full length BoNT/B.

A gene has also been assembled coding for a variant polypeptide, LH₄₁₇/B (SEQ ID NO: 22), which possesses an amino acid sequence at its carboxy terminus



equivalent by amino acid homology to that at the carboxy-terminus of the heavy chain fragment in native LH_N/A .

A gene has also been assembled coding for a variant polypeptide, LH_{107}/B (SEQ ID NO: 24), which expresses at its carboxy-terminus a short sequence from the amino terminus of the heavy chain of BoNT/B sufficient to maintain solubility of the expressed polypeptide.

Construct Variants

A variant of the coding sequence for the first 274 bases of the gene shown in SEQ ID NO: 21 has been produced which whilst being a non-native nucleotide sequence still codes for the native polypeptide.

Two double stranded, a 268 base pair and a 951 base pair, gene sequences have been created using an overlapping primer PCR strategy. The nucleotide bias of these sequences was designed to have an *E.coli* codon usage bias.

For the first sequence, six oligonucleotides representing the first (5') 268 nucleotides of the native sequence for botulinum toxin type B were synthesised. For the second sequence 23 oligonucleotides representing internal sequence nucleotides 691-1641 of the native sequence for botulinum toxin type B were synthesised. The oligonucleotides ranged from 57-73 nucleotides in length. Overlapping regions, 17-20 nucleotides, were designed to give melting temperatures in the range 52-56°C. In addition, terminal restriction endonuclease sites of the synthetic products were constructed to facilitate insertion of these products into the exact corresponding region of the native sequence. The 268 bp 5' synthetic sequence has been incorporated into the gene shown in SEQ ID NO: 21 in place of the original first 268 bases (and is shown in SEQ ID NO: 27). Similarly the sequence could be inserted into other genes of the examples.

Another-variant-sequence-equivalent-to-nucleotides-691-to-1641-of-SEQ-ID-NO:-21-

, and employing non-native codon usage whilst coding for a native polypeptide sequence, has been constructed using the internal synthetic sequence. This sequence (SEQ ID NO: 28) can be incorporated, alone or in combination with other variant sequences, in place of the equivalent coding sequence in any of the genes of the example.

Example 3

An exemplification of the utility of this invention is as a non-toxic and effective immunogen. The non-toxic nature of the recombinant, single chain material was demonstrated by intraperitoneal administration in mice of GST-2LH₄₂₃/A. The polypeptide was prepared and purified as described above. The amount of immunoreactive material in the final preparation was determined by enzyme linked immunosorbent assay (ELISA) using a monoclonal antibody (BA11) reactive against a conformation dependent epitope on the native LH_N/A. The recombinant material was serially diluted in phosphate buffered saline (PBS; NaCl 8 g/l, KCl 0.2 g/l, Na₂HPO₄ 1.15 g/l, KH₂PO₄ 0.2 g/l, pH 7.4) and 0.5 ml volumes injected into 3 groups of 4 mice such that each group of mice received 10, 5 and 1 micrograms of material respectively. Mice were observed for 4 days and no deaths were seen.

For immunisation, 20 μ g of GST-₂LH₄₂₃/A in a 1.0 ml volume of water-in-oil emulsion (1:1 vol:vol) using Freund's complete (primary injections only) or Freund's incomplete adjuvant was administered into guinea pigs via two sub-cutaneous dorsal injections. Three injections at 10 day intervals were given (day 1, day 10 and day 20) and antiserum collected on day 30. The antisera were shown by ELISA to be immunoreactive against native botulinum neurotoxin type A and to its derivative LH_N/A. Antisera which were botulinum neurotoxin reactive at a dilution of 1:2000 were used for evaluation of neutralising efficacy in mice. For neutralisation assays 0.1 ml of antiserum was diluted into 2.5 ml of gelatine phosphate buffer (GPB; Na₂HPO₄ anhydrous 10 g/l, gelatin (Difco) 2 g/l, pH 6.5-6.6) containing a dilution range from 0.5 μ g (5X10⁻⁶ g) to 5 picograms (5X10⁻¹² g). Aliquots of 0.5 ml wer injected into mice intraperitoneally and deaths recorded



over a 4 day period. The results are shown in Table 1 and Table 2. It can clearly be seen that 0.5 ml of 1:40 diluted anti- $GST_{-2}LH_{423}/A$ antiserum can protect mice against intraperitoneal challenge with botulinum neurotoxin in the range 5 pg - 50 ng (1 - 10,000 mouse LD50; 1 mouse LD50 = 5 pg).

TABLE 1. Neutralisation of botulinum neurotoxin in mice by guinea pig anti-GST-2LH₄₂₃/A antiserum.

		<u>B</u>	lotulinum Tox	in/mouse			
Survivors On Day	0.5 <i>µ</i> g	0.005µg	0.0005µg	0.5ng	0.005ng	5pg	Control (no toxin)
1	0	4	4	4	4	4	4
2	-	4	4	4	4	4	4
3	•	4	4	4	4	4	4
4	-	4	4	4	4	4	4

TABLE 2. Neutralisation of botulinum neurotoxin in mice by non-immune guin a pig antiserum.

		<u> </u>	otulinum Tox	in/mouse			
Survivors On Day	0.5µg	0.005µg	0.0005µg	0.5ng	0.005ng	5pg	Control (no toxin)
1	0	0	o	0	o	2	4
2	-	-	•	-	-	0	4
з.	•	-	•	•	-	-	4
4	-	•	•	-	-	-	4

Example 4

Expression of recombinant LH₁₀₇/B in *E. coli*.

As an exemplification of the expression of a nucleic acid coding for a LH_N of a clostridial neurotoxin of a serotype other than botulinum neurotoxin type A, the nucleic acid sequence (SEQ ID NO: 23) coding for the polypeptide LH_{107}/B (SEQ ID

NO: 24) was inserted into the commercially available plasmid pET28a (Novogen, Madison, WI, USA). The nucleic acid was expressed in $E.\ coli$ BL21 (DE3) (New England BioLabs, Beverley, MA, USA) as a fusion protein with a N-terminal T7 fusion peptide, under IPTG induction at 1 mM for 90 minutes at 37°C. Cultures were harvested and recombinant protein extracted as described previously for LH_{423}/A .

Recombinant protein was recovered and purified from bacterial paste lysates by immunoaffinity adsorption to an immobilised anti-T7 peptide monoclonal antibody using a T7 tag purification kit (New England bioLabs, Beverley, MA, USA). Purified recombinant protein was analysed by gradient (4-20%) denaturing SDS-polyacrylamide gel electrophoresis (Novex, San Diego, CA, USA) and western blotting using polyclonal anti-botulinum neurotoxin type antiserum or anti-T7 antiserum. Western blotting reagents were from Novex, immunostained proteins were visualised using the Enhanced Chemi-Luminescence system (ECL) from Amersham. The expression of an anti-T7 antibody and anti-botulinum neurotoxin type B antiserum reactive recombinant product is demonstrated in Figure 13.

The recombinant product was soluble and retained that part of the light chain responsible for endopeptidase activity.

The invention thus provides recombinant polypeptides useful inter alia as immunogens, enzyme standards and components for synthesis of molecules as described in WO-A-94/21300.

. ..

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT:
 - (A) NAME: MICROBIOLOGICAL RESEARCH AUTHORITY
 - (B) STREET: Centre For Applied Microbiology And Research, Porton Down
 - (C) CITY: Salisbury
 - (D) STATE: Wiltshire
 - (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): SP4 0JG
 - (A) NAME: THE SPEYWOOD LABORATORY LIMITED
 - (B) STREET: 14 Kensington Square
 - (C) CITY: London
 - (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): W8 5HH
 - (A) NAME: FOSTER; Keith Alan
 - (B) STREET: Centre For Applied Microbiology And Research, Porton Down

 - (C) CITY: Salisbury
 (D) STATE: Wiltshire
 - (E) COUNTRY: UK
 - (F) POSTAL CODE (ZIP): SP4 0JG
 - (A) NAME: QUINN; Conrad Padraig
 - (B) STREET: Centre For Applied Microbiology And Research, Porton Down

 - (C) CITY: Salisbury (D) STATE: Wiltshire (E) COUNTRY: UK

 - (F) POSTAL CODE (ZIP): SP4 0JG
 - (A) NAME: SHONE; Clifford Charles
 - (B) STREET: Centre For Applied Microbiology And Research, Porton Down

 - (C) CITY: Salisbury
 (D) STATE: Wiltshire
 - (E). COUNTRY: UK
 - (F) POSTAL CODE (ZIP): SP4 0JG
- (ii) TITLE OF INVENTION: Recombinant Toxin Fragments
- (iii) NUMBER OF SEQUENCES: 28
- (iv) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
- (2) INFORMATION FOR SEQ ID NO: 1:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2616 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)



(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:1..2616

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

ATG Met 1	CAG	TTC Phe	GTG Val	AAC Asn 5	AAG Lys	CAG Gln	TTC Phe	AAC Asn	TAT Tyr 10	AAG Lys	GAC Asp	CCT Pro	GTA Val	AAC Asn 15	GGT Gly		48
GTT Val	GAC Asp	ATT	GCC Ala 20	Tyr	ATC Ile	AAA Lys	ATT Ile	CCA Pro 25	Asn	GCC Ala	GGC Gly	CAG Gln	ATG Met 30	CAG Gln	CCG Pro		96
GTG Val	AAG Lys	GCT Ala 35	TTC Phe	AAG Lys	ATT Ile	CAT His	AAC Asn 40	AAA Lys	ATC Ile	TGG Trp	GTT Val	ATT Ile 45	CCG Pro	GAA Glu	CGC Arg		144
GAT Asp	ACA Thr 50	TTT Phe	ACG Thr	AAC Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	CCG Pro	CCG Pro	CCG Pro	GAA Glu		192
A1a 65	Lys	CAG Gln	Val	Pro	Val 70	Ser	Tyr	Tyr	Asp	Ser 75	Thr	Tyr	Leu	Ser	Thr 80	:	240
Asp	Asn	GAG Glu	Lys	Asp 85	Asn	Tyr	Leu	Lys	Gly 90	Val	Thr	Lys	Leu	Phe 95	Glu		288
Arg	Ile	TAT Tyr	Ser 100	Thr	Asp	Leu	Gly	Arg 105	Met	Leu	Leu	Thr	Ser 110	Ile	Val	;	336
Arg	Gly	ATC Ile 115	Pro	Phe	Trp	Gly	Gly 120	Ser	Thr	Ile	Asp	Thr 125	Glu	Leu	Lys	3	384
Val	Ile 130	GAC Asp	Thr	Asn	Cys	11e 135	Asn	Val	Ile	Gln	Pro 140	Asp	Gly	Ser	Tyr	4	132
Arg 145	Ser	GAA Glu	Glu	Leu	150	Leu	Val	Ile	Ile	Gly 155	Pro	Ser	Ala	Asp	Ile 160	4	180
Ile	Gln	TTT Phe	Glu	Cys 165	Lys	Ser	Phe	Gly	His 170	Glu	Val	Leu	Asn	Leu 175	Thr	S	28
Arg	Asn		Tyr 180	Gly	Ser	Thr	Gln	Tyr 185	Ile	Arg	Phe	Ser	Pro 190	Asp	Phe	S	76
Thr	Phe	GGT Gly 195	Phe	Glu	Glu	Ser	Leu 200	Glu	Val	Asp	Thr	Asn 205	Pro	Leu	Leu	6	24
Gly	Ala 210	GGC Gly	Lys	Phe	Ala	Thr 215	qaA	Pro	Ala	Val	Thr 220	Leu	Ala	His	Glu	6	72
Leu	Ile	CAC His	Ala	Gly	His	Arg	Leu	Tyr	Gly	Ile	Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	7	20

- 31 -

CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys	816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GTG Val	912
			GCT Ala													960
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
			AAG Lys 340													1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	T T T Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104
TTT Phe	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	1152
			GAT Asp													1200
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	1248
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
GCA Ala	TTA Leu 450	AAT Asn	GAT Asp	TTA Leu	TGT Cys	ATC Ile 455	AAA Lys	GTT Val	AAT Asn	AAT Asn	TGG Trp 460	GAC Asp	TTG Leu	TTT Phe	TTT Phe	1392
			GAA Glu													1440
ATT Ile	ACA Thr	TCT Ser	GAT Asp	ACT Thr 485	TAA neA	ATA Ile	GAA Glu	GCA Ala	GCA Ala 490	GAA Glu	GAA Glu	AAT Asn	ATT Ile	AGT Ser 495	TTA Leu	1488
GAT Asp	TTA Leu	ATA Ile	CAA Gln 500	CAA Gln	TAT Tyr	TAT Tyr	TTA Leu	ACC Thr 505	TTT Phe.	AAT Asn.	TTT Phe-	GAT -Asp-	AAT Asn- 510	GAA Glu	CCT Pro-	1536

					GAA Glu											1584
					ATA Ile											1632
TTA Leu 545	GAT Asp	AAA Lys	TAT Tyr	ACT Thr	ATG Met 550	TTC Phe	CAT His	TAT Tyr	CTT Leu	CGT Arg 555	GCT Ala	CAA Gln	GAA Glu	TTT Phe	GAA Glu 560	1680
CAT His	GGT Gly	AAA Lys	TCT Ser	AGG Arg 565	ATT Ile	GCT Ala	TTA Leu	ACA Thr	AAT Asn 570	TCT Ser	GTT Val	AAC Asn	GAA Glu	GCA Ala 575	TTA Leu	1728
TTA Leu	AAT Asn	CCT Pro	AGT Ser 580	CGT Arg	GTT Val	TAT Tyr	ACA Thr	TTT Phe 585	TTT Phe	TCT Ser	TCA Ser	GAC Asp	TAT Tyr 590	GTA Val	AAG Lys	1776
AAA Lys	GTT Val	AAT Asn 595	AAA Lys	GCT Ala	ACG Thr	GAG Glu	GCA Ala 600	GCT Ala	ATG Met	TTT Phe	TTA Leu	GGC Gly 605	TGG Trp	GTA Val	GAA Glu	1824
CAA Gln	TTA Leu 610	GTA Val	TAT Tyr	GAT Asp	TTT Phe	ACC Thr 615	GAT Asp	GAA Glu	ACT Thr	AGC Ser	GAA Glu 620	GTA Val	AGT Ser	ACT Thr	ACG Thr	1872
GAT Asp 625	AAA Lys	ATT Ile	GCG Ala	GAT Asp	ATA Ile 630	ACT Thr	ATA Ile	ATT Ile	ATT Ile	CCA Pro 635	TAT Tyr	ATA Ile	GGA Gly	CCT Pro	GCT Ala 640	1920
TTA Leu	AAT Asn	ATA Ile	GGT Gly	AAT Asn 645	ATG Met	TTA Leu	TAT Tyr	AAA Lys	GAT Asp 650	GAT Asp	TTT Phe	GTA Val	GGT Gly	GCT Ala 655	TTA Leu	1968
ATA Ile	TTT Phe	TCA Ser	GGA Gly 660	GCT Ala	GTT Val	ATT Ile	CTG Leu	TTA Leu 665	GAA Glu	TTT Phe	ATA Ile	CCA Pro	GAG Glu 670	ATT Ile	GCA Ala	2016
ATA Ile	CCT Pro	GTA Val 675	TTA Leu	GGT Gly	ACT Thr	TTT Phe	GCA Ala 680	CTT Leu	GTA Val	TCA Ser	TAT Tyr	ATT Ile 685	GCG Ala	AAT Asn	AAG Lys	2064
GTT Val	CTA Leu 690	ACC Thr	GTT Val	CAA Gln	ACA Thr	ATA Ile 695	GAT Asp	AAT Asn	GCT Ala	TTA Leu	AGT Ser 700	AAA Lys	AGA Arg	AAT Asn	GAA Glu	2112
AAA Lys 705	TGG Trp	GAT Asp	GAG Glu	GTC Val	TAT Tyr 710	AAA Lys	TAT Tyr	ATA Ile	GTA Val	ACA Thr 715	AAT Asn	TGG Trp	TTA Leu	GCA Ala	AAG Lys 720	2160
GTT Val	AAT Asn	ACA Thr	CAG Gln	ATT Ile 725	GAT Asp	CTA Leu	ATA Ile	AGA Arg	AAA Lys 730	AAA Lys	ATG Met	AAA Lys	GAA Glu	GCT Ala 735	TTA Leu	2208
GAA Glu	AAT Asn	CAA Gln	GCA Ala 740	GAA Glu	GCA Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT Tyr	CAG Gln 750	TAT Tyr	AAT Asn	2256
CAA Gln	TAT Tyr	ACT Thr 755	GAG Glu	GAA Glu	GAG Glu	AAA Lys	AAT Asn 760	Asn	ATT Ile	AAT Asn	TTT Phe	AAT Asn 765	ATT Ile	GAT Asp	GAT Asp	2304
TTA Leu	AGT Ser 770	TCG Ser	AAA Lys	CTT —Leu	TAA neA-	GAG -Glu 775	-Ser	ATA Ile	TAA naA-	AAA Lys	GCT -Ala 780	_Met	ATT _Ile	AAT Asn	ATA _Ile_	 2352

PCT/GB97/02273

- 33 -

		AAT Asn							2400
		GTT Val 805							2448
		AAG Lys							2496
		TTA Leu							2544
		CTT Leu							2592
		TAT Tyr		TAA *					2616

(2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 872 amino acids
 - (B) TYPE: amino acid(D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys

Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr

Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile

Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr-170

Arg	Asn	Gly	Tyr 180	Gly	Ser	Thr	Gln	Tyr 185	Ile	Arg	Phe	Ser	Pro 190	Asp	Phe
Thr	Phe	Gly 195	Phe	Glu	Glu	Ser	Leu 200	Glu	Val	Asp	Thr	Asn 205	Pro	Leu	Leu
_	Ala 210	Gly	Lys	Phe	Ala	Thr 215	Asp	Pro	Ala	Val	Thr 220	Leu	Ala	His	Glu
Leu 225	Ile	His	Ala	Gly	His 230	Arg	Leu	Tyr	Gly	Ile 235	Ala	Ile	Asn	Pro	Asn 240
Arg	Val	Phe	Lys	Val 245	Asn	Thr	Asn	Ala	Tyr 250	Tyr	Glu	Met	Ser	Gly 255	Leu
Glu	Val	Ser	Phe 260	Glu	Glu	Leu	Arg	Thr 265	Phe	Gly	Gly	His	Asp 270	Ala	Lys
Phe	Ile	Asp 275	Ser	Leu	Gln	Glu	Asn 280	Glu	Phe	Arg	Leu	Tyr 285	Tyr	Tyr	Asn
Lys	Phe 290	Lys	Asp	Ile	Ala	Ser 295	Thr	Leu	Asn	Lys	Ala 300	Lys	Ser	Ile	Val
Gly 305	Thr	Thr	Ala	Ser	Leu 310	Gln	Tyr	Met	Lys	Asn 315	Val	Phe	Lys	Glu	Lys 320
Tyr	Leu	Leu	Ser	Glu 325	Asp	Thr	Ser	Gly	Lys 330	Phe	Ser	Val	Asp	Lys 335	Leu
Lys	Phe	Asp	Lys 340	Leu	Tyr	Lys	Met	Leu 345	Thr	Glu	Ile	Tyr	Thr 350	Glu	Asp
Asn	Phe	Val 355	Lys	Phe	Phe	Lys	Val 360	Leu	Asn	Arg	Lys	Thr 365	Tyr	Leu	Asn
Phe	Asp	Lys	Ala	Val	Phe		Ile	Asn	Ile	Val	Pro 380	Lys	Val	Asn	Tyr
	370	•				375					300				
Thr 385	370	_		Gly	Phe 390		Leu	Arg	Asn	Thr 395		Leu	Ala	Ala	Asn 400
385	370 Ile	Tyr	Asp		390	Asn				395	Asn		`	Ala Lys 415	400
385 Phe	370 Ile Asn	Tyr Gly	Asp Gln	Asn 405	390 Thr	Asn Glu	Ile	Asn	Asn 410	395 Met	Asn Asn	Phe	Thr	Lys	400 Leu
385 Phe Lys	370 Ile Asn	Tyr Gly Phe	Asp Gln Thr 420	Asn 405 Gly	390 Thr Leu	Asn Glu Phe	Ile Glu	Asn Phe 425	Asn 410 Tyr	395 Met Lys	Asn Asn Leu	Phe Leu	Thr Cys 430	Lys 415	400 Leu Arg
385 Phe Lys Gly	370 Ile Asn Asn	Tyr Gly Phe Ile 435	Asp Gln Thr 420	Asn 405 Gly Ser	390 Thr Leu Lys	Asn Glu Phe Thr	Ile Glu Lys 440	Asn Phe 425 Ser	Asn 410 Tyr Leu	395 Met Lys Asp	Asn Asn Leu Lys	Phe Leu Gly 445	Thr Cys 430 Tyr	Lys 415 Val	400 Leu Arg Lys
385 Phe Lys Gly Ala	370 Ile Asn Asn Ile Leu 450	Tyr Gly Phe Ile 435 Asn	Asp Gln Thr 420 Thr	Asn 405 Gly Ser Leu	390 Thr Leu Lys Cys	Asn Glu Phe Thr Ile 455	Ile Glu Lys 440 Lys	Asn Phe 425 Ser Val	Asn 410 Tyr Leu Asn	395 Met Lys Asp Asn	Asn Asn Leu Lys Trp 460	Phe Leu Gly 445 Asp	Thr Cys 430 Tyr	Lys 415 Val Asn	400 Leu Arg Lys Phe
385 Phe Lys Gly Ala Ser 465	370 Ile Asn Asn Ile Leu 450 Pro	Tyr Gly Phe Ile 435 Asn Ser	Asp Gln Thr 420 Thr Glu	Asn 405 Gly Ser Leu Asp	390 Thr Leu Lys Cys Asn 470	Asn Glu Phe Thr Ile 455 Phe	Ile Glu Lys 440 Lys	Asn Phe 425 Ser Val Asn	Asn 410 Tyr Leu Asn	395 Met Lys Asp Asn Leu 475	Asn Leu Lys Trp 460 Asn	Phe Leu Gly 445 Asp	Thr Cys 430 Tyr Leu Gly	Lys 415 Val Asn	Arg Lys Phe Glu 480
Phe Lys Gly Ala Ser 465	370 Ile Asn . Asn Ile Leu 450 Pro	Tyr Gly Phe Ile 435 Asn Ser	Asp Gln Thr 420 Thr Glu Asp	Asn 405 Gly Ser Leu Asp Thr 485	390 Thr Leu Lys Cys Asn 470 Asn	Asn Glu Phe Thr Ile 455 Phe Ile	Ile Glu Lys 440 Lys Thr	Asn Phe 425 Ser Val Asn Ala	Asn 410 Tyr Leu Asn Asp	395 Met Lys Asp Asn Leu 475 Glu	Asn Leu Lys Trp 460 Asn Glu	Phe Leu Gly 445 Asp Lys	Thr Cys 430 Tyr Leu Gly	Lys 415 Val Asn Phe Glu . Ser	Arg Lys Phe Glu 480 Leu

Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu 535 540 Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 570 Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys 585 Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu 600 Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala 630 Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala 665 Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys 680 Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu 730 Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile 775 Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met 790 Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly 825 Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 855 Thr Phe Thr Glu Tyr Ile Lys

- -=::: :

^{(2) -} INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2685 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS (B) LOCATION:1..2685

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGA Gly 1	TCC Ser	CCA Pro	GGA Gly	ATT Ile 5	CAT His	ATG Met	ACG Thr	TCG Ser	ACG Thr 10	CGT Arg	CTG Leu	CAG Gln	AAG Lys	CTT Leu 15	CTA Leu	4	8
GAA Glu	TTC Phe	GAG Glu	CTC Leu 20	CCG Pro	GGT Gly	ACC Thr	ATG Met	GAG Glu 25	TTC Phe	GTG Val	AAC Asn	AAG Lys	CAG Gln 30	TTC Phe	AAC Asn	9	96
TAT Tyr	AAG Lys	GAC Asp 35	CCT Pro	GTA Val	AAC Asn	GGT Gly	GTT Val 40	GAC Asp	ATT Ile	GCC Ala	TAC Tyr	ATC Ile 45	AAA Lys	ATT Ile	CCA Pro	14	4
AAG Lys	TAC Tyr 50	GGC	CAG Gln	ATG Met	CAG Gln	CCG Pro 55	GTG Val	AAG Lys	GCT Ala	TTC Phe	AAG Lys 60	ATT Ile	CAT His	AAC Asn	AAA Lys	19	12
ATC Ile 65	TGG Trp	GTT Val	ATT Ile	CCG Pro	GAA Glu 70	CGC Arg	GAT Asp	ACA Thr	TTT Phe	ACG Thr 75	AAC Asn	CCG Pro	GAA Glu	GAA Glu	GGA Gly 80	24	0
GAC Asp	TTG Leu	AAC Asn	CCG Pro	CCG Pro 85	CCG Pro	GAA Glu	GCA Ala	AAG Lys	CAG Gln 90	GTG Val	CCA Pro	GTT Val	TCA Ser	TAC Tyr 95	TAC Tyr	28	8
GAT Asp	TCA Ser	ACC Thr	TAT Tyr 100	CTG Leu	AGC Ser	ACA Thr	GAC Asp	AAC Asn 105	GAG Glu	AAG Lys	GAT Asp	AAC Asn	TAC Tyr 110	CTG Leu	AAG Lys	33	6
GGA Gly	GTG Val	ACC Thr 115	AAA Lys	TTA Leu	TTC Phe	GAG Glu	CGT Arg 120	ATT Ile	TAT Tyr	TCC Ser	ACT Thr	GAC Asp 125	CTG Leu	GGC Gly	CGT Arg	38	4
ATG Met	CTG Leu 130	CTG Leu	ACC Thr	TCA Ser	ATC Ile	GTC Val 135	CGC Arg	GGA Gly	ATC Ile	CCA Pro	TTT Phe 140	TGG Trp	GGT Gly	GGC Gly	AGT Ser	43	2
ACC Thr 145	ATT Ile	GAC Asp	ACG Thr	GAG Glu	TTG Leu 150	AAG Lys	GTT Val	ATT Ile	GAC Asp	ACT Thr 155	AAC Asn	TGC Cys	ATT Ile	AAC Asn	GTG Val 160	48	0
ATC Ile	CAA Gln	CCA Pro	GAC Asp	GGT Gly 165	AGC Ser	TAC Tyr	AGA Arg	TCT Ser	GAA Glu 170	GAA Glu	CTT Leu	AAC Asn	CTC Leu	GTA Val 175	ATC Ile	52	8
					GAC Asp											57	6
	Glu		Leu		CTG Leu		Arg									62	4

ATT Ile	CGT Arg 210	Phe	AGC Ser	CCA Pro	GAC Asp	TTC Phe 215	Thr	TTC Phe	GGT	TTC Phe	GA0	ı Glı	G AGO	CTC	G GAG	672
GTT Val 225	Asp	ACC Thr	AAC Asn	CCG Pro	CTG Leu 230	Leu	GGT Gly	GCA Ala	GGC Gly	AAG Lys 235	Phe	GCA Ala	ACT Thr	GA1 Asp	CCA Pro 240	720
GCG Ala	GTG Val	ACC Thr	CTG Leu	GCA Ala 245	His	GAG Glu	CTG Leu	ATC Ile	CAC His 250	GCC Ala	GGT Gly	CAT His	CGT Arg	CTC Leu 255	TAT	768
GGC Gly	ATT Ile	GCG Ala	ATT Ile 260	AAC Asn	CCG Pro	AAC Asn	CGC Arg	GTG Val 265	TTC Phe	AAG Lys	GTT Val	AAC Asn	ACC Thr 270	Asn	GCC Ala	816
TAC Tyr	TAC Tyr	GAG Glu 275	ATG Met	AGT Ser	GGT Gly	TTA Leu	GAA Glu 280	GTA Val	AGC Ser	TTC Phe	GAG Glu	GAA Glu 285	Leu	CGC	ACG Thr	864
TTC Phe	GGT Gly 290	GGC Gly	CAT His	GAT Asp	GCG Ala	AAG Lys 295	TTT Phe	ATC Ile	GAC Asp	AGC Ser	TTG Leu 300	CAG Gln	GAG Glu	AAC Asn	GAG Glu	912
TTC Phe 305	CGT Arg	CTG Leu	TAC Tyr	TAC Tyr	TAC Tyr 310	AAC Asn	AAG Lys	TTT Phe	AAA Lys	GAT Asp 315	ATT Ile	GCA Ala	AGT Ser	ACA Thr	CTG Leu 320	960
AAC Asn	AAG Lys	GCT Ala	AAG Lys	TCC Ser 325	ATT Ile	GTG Val	GGT Gly	ACC Thr	ACT Thr 330	GCT Ala	TCA Ser	TTA Leu	CAG Gln	TAT Tyr 335	ATG Met	1008
AAA Lys	AAT Asn	GTT Val	TTT Phe 340	AAA Lys	GAG Glu	AAA Lys	TAT Tyr	CTC Leu 345	CTA Leu	TCT Ser	GAA Glu	GAT Asp	ACA Thr 350	TCT Ser	GGA Gly	1056
AAA Lys	TTT Phe	TCG Ser 355	GTA Val	GAT Asp	AAA Lys	TTA Leu	AAA Lys 360	TTT Phe	GAT Asp	AAG Lys	TTA Leu	TAC Tyr 365	AAA Lys	ATG Met	TTA Leu	1104
ACA Thr	GAG Glu 370	ATT Ile	TAC Tyr	ACA Thr	GAG Glu	GAT Asp 375	AAT Asn	TTT Phe	GTT Val	AAG Lys	TTT Phe 380	TTT Phe	AAA Lys	GTA Val	CTT Leu	1152
AAC Asn 385	AGA Arg	AAA Lys	ACA Thr	TAT Tyr	TTG Leu 390	AAT Asn	TTT Phe	GAT Asp	AAA Lys	GCC Ala 395	GTA Val	TTT Phe	AAG Lys	ATA Ile	AAT Asn 400	1200
ATA Ile	GTA Val	CCT Pro	AAG Lys	GTA Val 405	AAT Asn	TAC	ACA Thr	ATA Ile	TAT Tyr 410	GAT Asp	GGA Gly	TTT Phe	AAT Asn	TTA Leu 415	AGA Arg	1248
AAT Asn	ACA Thr	AAT Asn	TTA Leu 420	GCA Ala	GCA Ala	AAC Asn	TTT Phe	AAT Asn 425	GGT Gly	CAA Gln	AAT Asn	ACA Thr	GAA Glu 430	ATT Ile	AAT Asn	1296
AAT Asn	ATG Met	AAT Asn 435	TTT Phe	ACT Thr	AAA Lys	CTA Leu	AAA Lys 440	AAT Asn	TTT Phe	ACT Thr	GGA Gly	TTG Leu 445	TTT Phe	GAA Glu	TTT Phe	1344
TAT Tyr	AAG Lys 450	TTG Leu	CTA Leu	TGT Cys	GTA Val	AGA Arg 455	GGG	ATA Ile	ATA Ile	ACT Thr	TCT Ser 460	AAA Lys	ACT Thr	AAA Lys	TCA Ser	1392
TTA Leu 465	qaA	Lys	Gly	Tyr	Asn	AAG Lys	Ala	Leu	Asn	Asp	Leu	Cys	Ile	Lys	Val	1440

2 2222 2 22222 2

AAT Asn	AAT Asn	TGG Trp	GAC Asp	TTG Leu 485	TTT Phe	TTT Phe	AGT Ser	CCT Pro	TCA Ser 490	GAA Glu	GAT Asp	AAT Asn	TTT Phe	ACT Thr 495	AAT Asn	1488
GAT Asp	CTA Leu	AAT Asn	AAA Lys 500	GGA Gly	GAA Glu	GAA Glu	ATT Ile	ACA Thr 505	TCT Ser	GAT Asp	ACT Thr	AAT Asn	ATA Ile 510	GAA Glu	GCA Ala	1536
GCA Ala	GAA Glu	GAA Glu 515	AAT Asn	ATT Ile	AGT Ser	TTA Leu	GAT Asp 520	TTA Leu	ATA Ile	CAA Gln	CAA Gln	TAT Tyr 525	TAT Tyr	TTA Leu	ACC Thr	1584
					GAA Glu											1632
					CAA Gln 550											1680
					TAT Tyr											1728
					TTT Phe											1776
					GCA Ala											1824
					GTA Val											1872
					GTA Val 630											1920
					ACT Thr											1968
					CCT Pro											2016
					GCT Ala										TTA Leu	2064
					ATT Ile											2112
					AAT Asn 710											2160
					AAT Asn											2208
Val	Thr	Asn	TGG Trp 740	Leu	GCA Ala	AAG Lys	Val	Asn	Thr	CAG Gln	ATT	Asp	CTA Leu 750	Ile	AGA Arg	 2256

AAA Lys	AAA Lys	ATG Met 755	AAA Lys	GAA Glu	GCT Ala	TTA Leu	GAA Glu 760	AAT Asn	CAA Gln	GCA Ala	GAA Glu	GCA Ala 765	ACA Thr	AAG Lys	GCT Ala	2304
ATA Ile	ATA Ile 770	AAC Asn	TAT Tyr	CAG Gln	TAT Tyr	AAT Asn 775	CAA Gln	TAT Tyr	ACT Thr	GAG Glu	GAA Glu 780	GAG Glu	AAA Lys	AAT Asn	AAT Asn	23 52
	AAT Asn															2400
AAT Asn	AAA Lys	GCT Ala	ATG Met	ATT Ile 805	AAT Asn	ATA Ile	AAT Asn	AAA Lys	TTT Phe 810	TTG Leu	AAT Asn	CAA Gln	TGC Cys	TCT Ser 815	GTT Val	2448
TCA Ser	TAT Tyr	TTA Leu	ATG Met 820	AAT Asn	TCT Ser	ATG Met	ATC Ile	CCT Pro 825	TAT Tyr	GGT Gly	GTT Val	AAA Lys	CGG Arg 830	TTA Leu	GAA Glu	2496
GAT Asp	TTT Phe	GAT Asp 835	GCT Ala	AGT Ser	CTT Leu	AAA Lys	GAT Asp 840	GCA Ala	TTA Leu	TTA Leu	AAG Lys	TAT Tyr 845	ATA Ile	TAT Tyr	GAT Asp	2544
AAT Asn	AGA Arg 850	GGA Gly	ACT Thr	TTA Leu	ATT Ile	GGT Gly 855	CAA Gln	GTA Val	GAT Asp	AGA Arg	TTA Leu 860	AAA Lys	GAT Asp	AAA Lys	GTT Val	2592
AAT Asn 865	AAT Asn	ACA Thr	CTT Leu	AGT Ser	ACA Thr 870	GAT Asp	ATA Ile	CCT Pro	TTT Phe	CAG Gln 875	CTT Leu	TCC Ser	AAA Lys	TAC Tyr	GTA Val 880	2640
GAT Asp	AAT Asn	CAA Gln	AGA Arg	TTA Leu 885	TTA Leu	TCT Ser	ACA Thr	TTT Phe	ACT Thr 890	GAA Glu	TAT Tyr	ATT Ile	AAG Lys	TAA * 895		2685

(2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 895 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Gly Ser Pro Gly Ile His Met Thr Ser Thr Arg Leu Gln Lys Leu Leu Glu Phe Glu Leu Pro Gly Thr Met Glu Phe Val Asn Lys Gln Phe Asn 20 25 30 --

Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro 35 40 45

Lys Tyr Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys

Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly

Asp Leu Asn Pro Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr

Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys

Gly	Val	Thr 115	Lys	Leu	Phe	Glu	Arg 120	Ile	Tyr	Ser	Thr	Asp 125	Leu	Gly	Arg
Met	Leu 130	Leu	Thr	Ser	Ile	Val 135	Arg	Gly	Ile	Pro	Phe 140	Trp	Gly	Gly	Ser
Thr 145	Ile	Asp	Thr	Glu	Leu 150	Lys	Val	Ile	Asp	Thr 155	Asn	Cys	Ile	Asn	Val 160
Ile	Gln	Pro	Asp	Gly 165	Ser	Tyr	Arg	Ser	Glu 170	Glu	Leu	Asn	Leu	Val 175	Ile
Ile	Gly	Pro	Ser 180	Ala	Asp	Ile	Ile	Gln 185	Phe	Glu	Cys	Lys	Ser 190	Phe	Gly
His	Glu	Val 195	Leu	Asn	Leu	Thr	Arg 200	Asn	Gly	Tyr	Gly	Ser 205	Thr	Gln	Tyr
Ile	Arg 210	Phe	Ser	Pro	Asp	Phe 215	Thr	Phe	Gly	Phe	Glu 220	Glu	Ser	Leu	Glu
Val 225	Asp	Thr	Asn	Pro	Leu 230	Leu	Gly	Ala	Gly	Lys 235	Phe	Ala	Thr	Asp	Pro 240
Ala	Val	Thr	Leu	Ala 245	His	Glu	Leu	Ile	His 250	Ala	Gly	His	Arg	Leu 255	Tyr
Gly	Ile	Ala	11e 260	Asn	Pro	Asn	Arg	Val 265	Phe	Lys	Val	Asn	Thr 270	Asn	Ala
Tyr	Tyr	Glu 275	Met	Ser	Gly	Leu	Glu 280	Val	Ser	Phe	Glu	Glu 285	Leu	Arg	Thr
Phe	Gly 290	Gly	His	Asp	Ala	Lys 295	Phe	Ile	Asp	Ser	Leu 300	Gln	Glu	Asn	Glu
Phe 305	Arg	Leu	Tyr	Tyr	Tyr 310	Asn	Lys	Phe	Lys	Asp 315	Ile	Ala	Ser	Thr	Leu 320
	-		Lys	325					330					335	
Lys	Asn	Val	Phe 340	Lys	Glu	Lys	Tyr	Leu 345	Leu	Ser	Glu	Asp	Thr 350	Ser	Gly
		355	Val				360					365			
	370		Tyr			375					380				
385			Thr		390					395					400
Ile	Val	Pro	Lys	Val 405	Asn	Tyr	Thr	Ile	Tyr 410	Asp	Gly	Phe	Asn	Leu 415	Arg
			Leu 420					425					430		
Asn	Met	Asn 435	Phe	Thr	ГÀг	Leu	Lys 440	Asn	Phe	Thr	Gly	Leu 445	Phe	Glu	Phe
Tyr	Lys 450	Leu	Leu	Cys	Val	Arg 455	Gly	Ile	Ile	Thr	Ser 460	Lys	Thr	Lys	Ser

DUDGOOD 110 000700444

Leu	Asp	Lys	Gly	Tyr	Asn	Lys	Ala	Leu	Asn	Asp	Leu	Cys	Ile	Lys	Val
465	_				470					475,					480
Asn	Asn	Trp	Asp	Leu 485	Phe	Phe	Ser	Pro	Ser 490	Glu	Asp	Asn	Pne	495	Asn
Asp	Leu	Asn	Lys 500	Gly	Glu	Glu	Ile	Thr 505	Ser	Asp	Thr	Asn	Ile 510	Glu	Ala
Ala	Glu	Glu 515	Asn	Ile	Ser	Leu	Asp 520	Leu	Ile	Gln	Gln	Tyr 525	Tyr	Leu	Thr
Phe	Asn 530	Phe	Asp	Asn	Glu	Pro 535	Glu	Asn	Ile	Ser	Ile 540	Glu	Asn	Leu	Ser
Ser 545	Asp	Ile	Ile	Gly	Gln 550	Leu	Glu	Leu	Met	Pro 555	Asn	Ile	Glu	Arg	Phe 560
Pro	Asn	Gly	Lys	Lys 565	Tyr	Glu	Leu	Asp	Lys 570	Tyr	Thr	Met	Phe	His 575	Tyr
Leu	Arg	Ala	Gln 580	Glu	Phe	Glu	His	Gly 505	Lys	Ser	Arg	Ile	Ala 590	Leu	Thr
Asn	Ser	Val 5 95	Asn	Glu	Ala	Leu	Leu 600	Asn	Pro	Ser	Arg	Val 605	Tyr	Thr	Phe
Phe	Ser 610	Ser	Asp	Tyr	Val	Lys 615	Lys	Val	Asn	Lys	Ala 620	Thr	Glu	Ala	Ala
Met 625	Phe	Leu	Gly	Trp	Val 630	Glu	Gln	Leu	Val	Tyr 635	Asp	Phe	Thr	Asp	Glu 640
				645	Thr				650					655	
			660		Pro			665					670		
-	_	675			Ala		680					685			
	690				Ile	695					700				
705					Asn 710					715					720
Ala	Leu	Ser	Lys	Arg 725	Asn	Glu	Lys	Trp	Asp 730	Glu	Val	Tyr	Lys	Tyr 735	Ile
			740		Ala			745					750		
		755			Ala		760					765			
Ile	11e 770	Asn	Tyr	Gln	Tyr	Asn 775	Gln	Tyr	Thr	Glu	Glu 780	Glu	Lys	Asn	Asn
Ile 785	Asn	Phe	Asn	Ile	Asp 790	Asp	Leu	Ser	Ser	Lys 7 9 5	Leu	Asn	Glu	Ser	Ile 800
Asn	Lys	Ala	Met	11e 805	Asn	Ile	Asn	Lys	Phe 810	Leu	neA	Gln	Суз	Ser 815	Val

WO 98/07864 PCT/GB97/02273

- 42 -

Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 825 830 Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys * 890

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2622 base pairs (B) TYPE: nucleic acid

 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..2622
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

(GGA Gly 1	TCC Ser	ATG Met	GAG Glu	TTC Phe 5	GTG Val	AAC Asn	AAG Lys	CAG Gln	TTC Phe 10	AAC Asn	TAT Tyr	AAG Lys	GAC Asp	CCT Pro 15	GTA Val	48
1	AAC Asn	GGT Gly	GTT Val	GAC Asp 20	ATT Ile	GCC Ala	TAC Tyr	ATC Ile	AAA Lys 25	ATT Ile	CCA Pro	AAG Lys	TAC Tyr	GGC Gly 30	CAG Gln	ATG Met	96
(CAG Gln	CCG Pro	GTG Val 35	AAG Lys	GCT Ala	TTC Phe	AAG Lys	ATT Ile 40	CAT His	AAC Asn	AAA Lys	ATC Ile	TGG Trp 45	GTT Val	ATT Ile	CCG Pro	144
(GAA Glu	CGC Arg 50	GAT Asp	ACA Thr	TTT Phe	ACG Thr	AAC Asn 55	CCG Pro	GAA Glu	GAA Glu	GGA Gly	GAC Asp 60	TTG Leu	AAC Asn	CCG Pro	CCG Pro	192
1	CCG Pro 65	GAA Glu	GCA Ala	AAG Lys	CAG Gln	GTG Val 70	CCA Pro	GTT Val	TCA Ser	TAC Tyr	TAC Tyr 75	GAT Asp	TCA Ser	ACC Thr	TAT Tyr	CTG Leu 80	240
					GAG Glu 85												288
					TAT Tyr												336
					ATC Ile												384
		Lys			GAC Asp							Ile					432
		130					772					140					

- 43 -

	TAC Tyr															480
GAC Asp	ATT Ile	ATC Ile	CAG Gln	TTT Phe 165	GAG Glu	TGC Cys	AAG Lys	AGC Ser	TTT Phe 170	GGC Gly	CAC His	GAA Glu	GTG Val	TTG Leu 175	AAC Asn	528
	ACG Thr															576
GAC Asp	TTC Phe	ACG Thr 195	TTC Phe	GGT Gly	TTC Phe	GAG Glu	GAG Glu 200	AGC Ser	CTG Leu	GAG Glu	GTT Val	GAT Asp 205	ACC Thr	AAC Asn	CCG Pro	624
CTG Leu	TTG Leu 210	GGT Gly	GCA Ala	GGC Gly	AAG Lys	TTC Phe 215	GCA Ala	ACT Thr	GAT Asp	CCA Pro	GCG Ala 220	GTG Val	ACC Thr	CTG Leu	GCA Ala	672
	GAG Glu															720
CCG Pro	AAC Asn	CGC Arg	GTG Val	TTC Phe 245	AAG Lys	GTT Val	AAC Asn	ACC Thr	AAC Asn 250	GCC Ala	TAC Tyr	TAC Tyr	GAG Glu	ATG Met 255	AGT Ser	768
GGT Gly	TTA Leu	GAA Glu	GTA Val 260	AGC Ser	TTC Phe	GAG Glu	GAA Glu	CTG Leu 265	CGC Arg	ACG Thr	TTC Phe	GGT Gly	GGC Gly 270	CAT His	GAT Asp	816
GCG Ala	AAG Lys	TTT Phe 275	ATC Ile	GAC Asp	AGC Ser	TTG Leu	CAG Gln 280	GAG Glu	AAC Asn	GAG Glu	TTC Phe	CGT Arg 285	CTG Leu	TAC Tyr	TAC Tyr	864
TAC Tyr	AAC Asn 290	AAG Lys	TTT Phe	AAA Lys	GAT Asp	ATT Ile 295	GCA Ala	AGT Ser	ACA Thr	CTG Leu	AAC Asn 300	AAG Lys	GCT Ala	AAG Lys	TCC Ser	912
ATT Ile 305	GTG Val	GGT Gly	ACC Thr	ACT Thr	GCT Ala 310	TCA Ser	TTA Leu	CAG Gln	TAT Tyr	ATG Met 315	AAA Lys	AAT Asn	GTT Val	TTT Phe	AAA Lys 320	960
GAG Glu	AAA Lys	TAT Tyr	CTC Leu	CTA Leu 325	TCT Ser	GAA Glu	GAT Asp	ACA Thr	TCT Ser 330	GGA Gly	AAA Lys	TTT Phe	TCG Ser	GTA Val 335	GAT Asp	1008
AAA Lys	TTA Leu	AAA Lys	TTT Phe 340	GAT Asp	AAG Lys	TTA Leu	TAC Tyr	AAA Lys 345	ATG Met	TTA Leu	ACA Thr	GAG Glu	ATT Ile 350	TAC Tyr	ACA Thr	1056
GAG Glu	GAT Asp	AAT Asn 355	TTT Phe	GTT Val	AAG Lys	TTT Phe	TTT Phe 360	AAA Lys	GTA Val	CTT Leu	AAC Asn	AGA Arg 365	AAA Lys	ACA Thr	TAT Tyr	1104
TTG Leu	AAT Asn 370	TTT Phe	GAT Asp	AAA Lys	GCC Ala	GTA Val 375	TTT Phe	AAG Lys	ATA Ile	AAT Asn	ATA Ile 380	GTA Val	CCT Pro	AAG Lys	GTA Val	1152
AAT Asn 385	TAC Tyr	ACA Thr	ATA	TAT Tyr	GAT Asp 390	GGA Gly	TTT Phe	AAT Asn	TTA Leu	AGA Arg 395	AAT Asn	ACA Thr	AAT Asn	TTA Leu	GCA Ala 400	1200
GCA Ala	AAC Asn	Phe	Asn	GGT Gly -405	Gln	AAT Asn	Thr	Glu	Ile	Asn	Asn	ATG Met	Asn	Phe	ACT Thr	1248

										- 44	-					
AAA Lys	CTA Leu	AAA Lys	AAT Asn 420	TTT Phe	ACT Thr	GGA Gly	TTG Leu	TTT Phe 425	GAA Glu	TTT Phe	TAT Tyr	AAG Lys	TTG Leu 430	CTA Leu	TGT Cys	1296
GTA Val	AGA Arg	GGG Gly 435	ATA Ile	ATA Ile	ACT Thr	TCT Ser	AAA Lys 440	ACT Thr	AAA Lys	TCA Ser	TTA Leu	GAT Asp 445	AAA Lys	GGA Gly	TAC Tyr	1344
AAT Asn	AAG Lys 450	GCA Ala	TTA Leu	AAT Asn	GAT Asp	TTA Leu 455	TGT Cys	ATC Ile	AAA Lys	GTT Val	AAT Asn 460	AAT Asn	TGG Trp	GAC Asp	TTG Leu	1392
Phe 465	Phe	Ser	Pro	Ser	Glu 470	Asp	Asn	TTT Phe	Thr	Asn 475	Asp	Leu	Asn	Lys	Gly 480	1440
GAA Glu	GAA Glu	ATT Ile	ACA Thr	TCT Ser 485	GAT Asp	ACT Thr	AAT Asn	ATA Ile	GAA Glu 490	GCA Ala	GCA Ala	GAA Glu	GAA Glu	AAT Asn 495	ATT Ile	1488
Ser	Leu	Asp	Leu 500	Ile	Gln	Gln	Tyr	TAT Tyr 505	Leu	Thr	Phe	Asn	Phe 510	Asp	Asn	1536
Glu	Pro	Glu 515	Asn	Ile	Ser	Ile	Glu 520	AAT Asn	Leu	Ser	Ser	Asp 525	Ile	Ile	Gly	1584
Gln	Leu 530	Glu	Leu	Met	Pro	Asn 535	Ile	GAA Glu	Arg	Phe	Pro 540	Asn	Gly	Lys	Lys	1632
Tyr 545	Glu	Leu	Asp	Lys	Tyr 550	Thr	Met	TTC Phe	His	Tyr 555	Leu	Arg	Ala	Gln	Glu 560	1680
Phe	Glu	His	Gly	Lys 565	Ser	Arg	Ile	GCT Ala	Leu 570	Thr	Asn	Ser	Val	Asn 575	Glu	1728
Ala	Leu	Leu	Asn 580	Pro	Ser	Arg	Val	TAT Tyr 585	Thr	Phe	Phe	Ser	Ser 590	Asp	Tyr	1776
Val	Lys	Lys 595	Val	Asn	Lys	Ala	Thr 600	GAG Glu	Ala	Ala	Met	Phe 605	Leu	Gly	Trp	1824
Val	Glu 610	Gln	Leu	Val	Tyr	Asp 615	Phe	ACC Thr	Asp	Glu	Thr 620	Ser	Glu	Val	Ser	1872
Thr 625	Thr	Asp	Lys	Ile	Ala 630	Asp	Ile	ACT Thr	Ile	11e 635	Ile	Pro	Tyr	Ile	Gly 640	1920
Pro	Ala	Leu	Asn	11e 645	Gly	Asn	Met	TTA Leu	Tyr 650	Lys	Asp	Asp	Phe	Val 655	Gly	1968
GCT Ala	TTA Leu	ATA Ile	TTT Phe 660	TCA Ser	GGA Gly	GCT Ala	GTT Val	ATT Ile 665	CTG Leu	TTA Leu	GAA Glu	TTT Phe	ATA Ile 670	CCA Pro	GAG Glu	2016
ATT Ile_	GCA Ala_	ATA Ile 675	CCT Pro	GTA Val	TTA Leu	GGT Gly	ACT Thr 680	TTT Phe	GCA Ala	CTT Leu	GTA Val	TCA Ser 685	TAT Tyr	ATT Ile	GCG Ala	 2064

- 45 -

AAT Asn	AAG Lys 690	GTT Val	CTA Leu	ACC Thr	GTT Val	CAA Gln 695	ACA Thr	ATA Ile	GAT Asp	AAT Asn	GCT Ala 700	TTA Leu	AGT Ser	AAA Lys	AGA Arg	2112
AAT Asn 705	GAA Glu	AAA Lys	TGG Trp	GAT Asp	GAG Glu 710	GTC Val	TAT Tyr	AAA Lys	TAT	ATA Ile 715	GTA Val	ACA Thr	AAT Asn	TGG Trp	TTA Leu 720	2160
GCA Ala	AAG Lys	GTT Val	AAT Asn	ACA Thr 725	CAG Gln	ATT Ile	GAT Asp	CTA Leu	ATA Ile 730	AGA Arg	AAA Lys	AAA Lys	ATG Met	AAA Lys 735	GAA Glu	2208
GCT Ala	TTA Leu	GAA Glu	AAT Asn 740	CAA Gln	GCA Ala	GAA Glu	GCA Ala	ACA Thr 745	AAG Lys	GCT Ala	ATA Ile	ATA Ile	AAC Asn 750	TAT Tyr	CAG Gln	2256
TAT Tyr	AAT Asn	CAA Gln 755	TAT Tyr	ACT Thr	GAG Glu	GAA Glu	GAG Glu 760	AAA Lys	AAT Asn	AAT Asn	ATT Ile	AAT Asn 765	TTT Phe	AAT Asn	ATT Ile	2304
GAT Asp	GAT Asp 770	TTA Leu	AGT Ser	TCG Ser	AAA Lys	CTT Leu 775	AAT Asn	GAG Glu	TCT Ser	ATA Ile	AAT Asn 780	AAA Lys	GCT Ala	ATG Met	ATT Ile	2352
AAT Asn 785	ATA Ile	AAT Asn	AAA Lys	TTT Phe	TTG Leu 790	AAT Asn	CAA Gln	TGC Cys	TCT Ser	GTT Val 795	TCA Ser	TAT Tyr	TTA Leu	ATG Met	AAT Asn 800	2400
TCT Ser	ATG Met	ATC Ile	CCT Pro	TAT Tyr 805	GGT Gly	GTT Val	AAA Lys	CGG Arg	TTA Leu 810	GAA Glu	GAT Asp	TTT Phe	GAT Asp	GCT Ala 815	AGT Ser	2448
CTT Leu	AAA Lys	GAT Asp	GCA Ala 820	TTA Leu	TTA Leu	AAG Lys	TAT Tyr	ATA Ile 825	TAT Tyr	GAT Asp	AAT Asn	AGA Arg	GGA Gly 830	ACT Thr	TTA Leu	2496
ATT Ile	GGT Gly	CAA Gln 835	GTA Val	GAT Asp	AGA Arg	TTA Leu	AAA Lys 840	GAT Asp	AAA Lys	GTT Val	AAT Asn	AAT Asn 845	ACA Thr	CTT Leu	AGT Ser	2544
ACA Thr	GAT Asp 850	ATA Ile	CCT Pro	TTT Phe	CAG Gln	CTT Leu 855	TCC Ser	AAA Lys	TAC Tyr	GTA Val	GAT Asp 860	AAT Asn	CAA Gln	AGA Arg	TTA Leu	2592
				Thr		TAT Tyr	_		TAA *							2622

(2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 874 amino acids
 - (B) TYPE: amino acid (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

Gly Ser Met Glu Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val 1 5 10 15

Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Lys Tyr Gly Gln Met 20 25 30

Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro---- 45

Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro 55 Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala 150 Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro 185 Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro 200 Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser 245 250 Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp 265 Ala Lys Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser 295 Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys-Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr 345 Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val 375 Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala 395

Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr 410 Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys 425 Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly 470 Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile 485 490 Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly 520 Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala 680 Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu 730 Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln 745

Tyr	Asn	Gln 755	Tyr	Thr	Glu	Glu	Glu 760	Lys	Asn	Asn	Ile	Asn 765		Asn	Ile	
Asp	Asp 770	Leu	Ser	Ser	Lys	Leu 775	Asn	Glu	Ser	Ile	Asn 780	Lys	Ala	Met	Ile	
Asn 785	Ile	Asn	Lys	Phe	Leu 790	Asn	Gln	Cys	Ser	Val 795	Ser	Tyr	Leu	Met	Asn 800	
Ser	Met	Ile	Pro	Tyr 805	Gly	Val	Lys	Arg	Leu 810	Glu	Asp	Phe	Asp	Ala 815	Ser	
Leu	Lys	Asp	Ala 820	Leu	Leu	Lys	Tyr	Ile 825	Tyr	Asp	Asn	Arg	Gly 830	Thr	Leu	
Ile	Gly	Gln 835	Val	Asp	Arg	Leu	Lys 840	Asp	Lys	Val	Asn	Asn 845	Thr	Leu	Ser	
Thr	Asp 850	Ile	Pro	Phe	Gln	Leu 855	Ser	Lys	Tyr	Val	Asp 860	Asn	Gln	Arg	Leu	
Leu 865	Ser	Thr	Phe	Thr	Glu 870	Tyr	Ile	Lys	*							
(2)	INFO	ORMA:	rion	FOR	SEQ	ID N	10: 5	7:								
	(i)	() () ()	QUENC A) LE B) TY C) ST O) TO	ENGTH PE: TRANI	i: 26 nucl	513 b leic ESS:	acio doub	pai:	cs							
	(ii)	MOI	LECUI	E T	PE:	DNA	(ger	nomi	2)							
		FEA (A	LECUI ATURE A) NA B) LO	S: ME/I	ŒY:	CDS	_	nomio	=)		•					
	(ix)	FE! (!	ATURE A) NA	: ME/I CATI	ŒY: ION:1	CDS	313): 7:	•					
ATG Met 1	(ix)	FE# (# (E	ATURE A) NA B) LC	E: ME/F DCATI CE DE	ŒY: ION:1 ESCRI	CDS 26 IPTIC	513 ON: S	SEQ I	D NO	AAA	GAT	CCT Pro	GTA Val	AAT Asn 15	GGT Gly	48
Met 1 GTT	(ix) (xi) CCA Pro	FEA (A SEC TTT Phe	ATURE A) NA B) LO QUENO	C: AME/F CCATI CE DE AAT ASN 5	ŒY: ION:1 ESCRI AAA Lys	CDS 26 CPTIC CAA Gln	ON: STTT Phe	SEQ J AAT ASN	TAT Tyr 10	AAA Lys GCA	GAT Asp GGA	Pro	Val	Asn 15 CAA	Gly	48
Met 1 GTT Val	(ix) (xi) CCA Pro GAT Asp	FEA (A (E SEC TTT Phe ATT Ile	ATURE A) NA B) LO QUENO GTT Val GCT Ala	E: AME/F DCATI TE DE AAT ASN 5 TAT TYT	CEY: CON:1 CSCRI AAA Lys ATA Ile	CDS 26 CPTIC CAA Gln AAA Lys	ON: STTT Phe ATT Ile	SEQ J AAT Asn CCA Pro 25	TAT Tyr 10 AAT Asn	AAA Lys GCA Ala TGG	GAT Asp GGA Gly	Pro CAA Gln	Val ATG Met 30	Asn 15 CAA Gln	Gly CCA Pro	
Met 1 GTT Val GTA Val	(ix) (xi) CCA Pro GAT Asp AAA Lys	FEA (A SEC TTT Phe ATT Ile GCT Ala 35	ATURE A) NA B) LO QUENC GTT Val GCT Ala 20	E: AME/FOCATION CE DE AAT ASN 5 TAT TYT AAA Lys	CEY: ION:1 ESCRI AAA Lys ATA Ile ATT Ile	CDS 26 CPTIC CAA Gln AAA Lys CAT His	ON: STTT Phe ATT Ile AAT ASn 40 GAA	SEQ I AAT ASN CCA Pro 25 AAA Lys	TAT Tyr 10 AAT ASN ATA 11e	AAA Lys GCA Ala TGG Trp	GAT Asp GGA Gly GTT Val	Pro CAA Gln ATT Ile 45 CCA	Val ATG Met 30 CCA Pro	Asn 15 CAA Gln GAA Glu	Gly CCA Pro AGA Arg	96
GTT Val GTA Val GAT Asp	(ix) (xi) CCA Pro GAT Asp AAA Lys ACA Thr 50 AAA	SEC TTT Phe ATT Ile GCT Ala 35 TTT Phe	ATURE A) NA B) LO QUENO GTT Val GCT Ala 20 TTT Phe	E: AME/IDCATI CE DE AAT ASN 5 TAT Tyr AAA Lys AAT ASN	CEY: CON:1 ESCRI AAA Lys ATA Ile ATT Ile CCT Pro	CDS 26 CPTIC CAA Gln AAA Lys CAT His GAA Glu 55	ON: STTT Phe ATT Ile AAT ASD GAA Glu	SEQ I AAT ASN CCA Pro 25 AAA Lys GGA Gly	TAT Tyr 10 AAT Asn ATA Ile GAT Asp	AAA Lys GCA Ala TGG Trp TTA Leu	GAT Asp GGA Gly GTT Val AAT Asn 60 ACA	Pro CAA Gln ATT Ile 45 CCA Pro	Val ATG Met 30 CCA Pro CCA Pro	Asn 15 CAA Gln GAA Glu CCA Pro	CCA Pro AGA Arg GAA Glu	96 144

										- 49	•					
AGA Arg	ATT Ile	TAT	TCA Ser 100	ACT Thr	GAT Asp	CTT Leu	GGA Gly	AGA Arg 105	ATG Met	TTG Leu	TTA Leu	ACA Thr	TCA Ser 110	Ile	GTA Val	. 336
AGG Arg	GGA Gly	ATA Ile 115	CCA Pro	TTT Phe	TGG Trp	GGT Gly	GGA Gly 120	AGT Ser	ACA Thr	ATA Ile	GAT Asp	ACA Thr 125	GAA Glu	TTA Leu	AAA Lys	384
GTT Val	ATT Ile 130	GAT Asp	ACT Thr	AAT Asn	TGT Cys	ATT Ile 135	AAT Asn	GTG Val	ATA Ile	CAA Gln	CCA Pro 140	GAT Asp	GGT Gly	AGT Ser	TAT Tyr	432
AGA Arg 145	TCA Ser	GAA Glu	GAA Glu	CTT Leu	AAT Asn 150	CTA Leu	GTA Val	ATA Ile	ATA Ile	GGA Gly 155	CCC Pro	TCA Ser	GCT Ala	GAT Asp	ATT Ile 160	480
ATA Ile	CAG Gln	TTT Phe	GAA Glu	TGT Cys 165	AAA Lys	AGC Ser	TTT Phe	GGA Gly	CAT His 170	GAA Glu	GTT Val	TTG Leu	AAT Asn	CTT Leu 175	ACG Thr	528
CGA Arg	AAT Asn	GGT Gly	TAT Tyr 180	GGC Gly	TCT Ser	ACT Thr	CAA Gln	TAC Tyr 185	ATT Ile	AGA Arg	TTT Phe	AGC Ser	CCA Pro 190	GAT Asp	TTT Phe	576
ACA Thr	TTT	GGT Gly 195	TTT Phe	GAG Glu	GAG Glu	TCA Ser	CTT Leu 200	GAA Glu	GTT Val	GAT Asp	ACA Thr	AAT Asn 205	CCT Pro	CTT Leu	TTA Leu	624
GGT Gly	GCA Ala 210	GGC Gly	AAA Lys	TTT Phe	GCT Ala	ACA Thr 215	GAT Asp	CCA Pro	GCA Ala	GTA Val	ACA Thr 220	TTA Leu	GCA Ala	CAT His	GAA Glu	672
CTT Leu 225	ATA Ile	CAT His	GCT Ala	GGA Gly	CAT His 230	AGA Arg	TTA Leu	TAT Tyr	GGA Gly	ATA Ile 235	GCA Ala	ATT Ile	AAT Asn	CCA Pro	AAT Asn 240	720
Arg	Val	Phe	Lys	Val 245	Asn	ACT Thr	Asn	Ala	Tyr 250	Tyr	Glu	Met	Ser	Gly 255	Leu	768
Glu	Val	Ser	Phe 260	Glu	Glu	CTT Leu	Arg	Thr 265	Phe	Gly	Gly	His	Asp 270	Ala	Lys	816
TTT Phe	ATA Ile	GAT Asp 275	AGT Ser	TTA Leu	CAG Gln	GAA Glu	AAC Asn 280	GAA Glu	TTT Phe	CGT Arg	CTA Leu	TAT Tyr 285	TAT Tyr	TAT Tyr	AAT Asn	864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATA Ile	GCA Ala	AGT Ser 295	ACA Thr	CTT Leu	AAT Asn	AAA Lys	GCT Ala 300	AAA Lys	TCA Ser	ATA Ile	GTA Val	912
GGT Gly 305	ACT Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
						ACA Thr										1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
AAT Asn	TTT Phe	Val	Lys	Phe	Phe	AAA Lys	Val	Leu	Asn	Arg	AAA Lys	Thr	TAT Tyr	TTG Leu	AAT Asn	1104

TTT Phe	GAT Asp	, rae	GCC Ala	GTA Val	TTT Phe	AAG Lys	TTE	AAT Asn	ATA	GTA Val	Pro	Lys	GTA Val	AA1 Asn	TAC Tyr	1152
ACA Thr 385	ATA Ile	TAT	GAT Asp	GGA Gly	TTT Phe	' AAT Asn	TTA	AGA Arg	AAT Asn	ACA Thr	Asn	מידים י	GCA Ala	GCA Ala	AAC Asn	1200
TTT	AAT	GGT Gly	CAA Gln	AAT Asn 405	Inr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ÀΤC	יימב:	TTT Phe	ACT Thr	AAA Lys 415	Leu	1248
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	CTA	202	1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC	AAT Asn	AAG Lys	1344
GCA Ala	TTA Leu 450	AAT Asn	GAT Asp	TTA Leu	TGT Cys	ATC Ile 455	AAA Lys	GTT Val	AAT Asn	AAT Asn	TGG Trp 460	GAC Asp	TTG Leu	TTT Phe	TTT Phe	1392
AGT Ser 465	CCT Pro	TCA Ser	GAA Glu	GAT Asp	AAT Asn 470	TTT Phe	ACT Thr	AAT Asn	GAT Asp	CTA Leu 475	AAT Asn	AAA Lys	GGA Gly	GAA Glu	GAA Glu 480	1440
ATT Ile	ACA Thr	TCT Ser	GAT Asp	ACT Thr 485	AAT Asn	ATA Ile	GAA Glu	GCA Ala	GCA Ala 490	GAA Glu	GAA Glu	AAT Asn	ATT Ile	AGT Ser 495	TTA Leu	1488
GAT Asp	TTA Leu	ATA Ile	CAA Gln 500	CAA Gln	TAT Tyr	TAT Tyr	TTA Leu	ACC Thr 505	TTT Phe	AAT Asn	TTT Phe	GAT Asp	AAT Asn 510	GAA Glu	CCT Pro	1536
GAA Glu	AAT Asn	ATT Ile 515	TCA Ser	ATA Ile	GAA Glu	AAT Asn	CTT Leu 520	TCA Ser	AGT Ser	GAC Asp	ATT Ile	ATA Ile 525	GGC Gly	CAA Gln	TTA Leu	1584
GAA Glu	CTT Leu 530	ATG Met	CCT Pro	AAT Asn	ATA Ile	GAA Glu 535	AGA Arg	TTT Phe	CCT Pro	AAT Asn	GGA Gly 540	AAA Lys	AAG Lys	TAT Tyr	GAG Glu	1632
TTA Leu 545	Asp	AAA Lys	TAT Tyr	ACT Thr	Met	TTC Phe	CAT His	TAT Tyr	Leu	CGT Arg 555	GCT Ala	CAA Gln	GAA Glu	TTT Phe	GAA Glu 560	1680
CAT His	GGT Gly	AAA Lys	TCT Ser	AGG Arg 565	ATT Ile	GCT Ala	TTA Leu	ACA Thr	AAT Asn 570	TCT Ser	GTT Val	AAC Asn	GAA Glu	GCA Ala 525	TTA Leu	1728
TTA Leu	AAT Asn	CCT Pro	AGT Ser 580	CGT Arg	GTT Val	TAT Tyr	ACA Thr	TTT Phe 585	TTT Phe	TCT Ser	TCA Ser	GAC Asp	TAT Tyr 590	GTA Val	AAG Lys	1776
AAA Lys	GTT Val	AAT Asn 595	AAA Lys	GCT Ala	ACG Thr	Glu	GCA Ala 600	GCT Ala	ATG Met	TTT Phe	TTA Leu	GGC Gly 605	TGG Trp	GTA Val	GAA Glu	1824
CAA Gln	TTA Leu 610	GTA Val	TAT Tyr	GAT Asp	Phe	ACC Thr 615	GAT Asp	GAA Glu	ACT Thr	AGC Ser	GAA Glu 620	GTA Val	AGT Ser	ACT Thr	ACG Thr	1872
GAT Asp 625	AAA Tys	ATT Ile	GCG Ala	GAT Asp	ATA Ile 630	ACT Thr	ATA Ile-	ATT Tie-	ATT The	CCA Pro- 635	TAT Tyr-	ATA Ile	GGA Gly-	CCT Pro	GCT Ala 640	1920

- 51 -

					ATG Met											1968
ATA Ile	TTT Phe	TCA Ser	GGA Gly 660	GCT Ala	GTT Val	ATT Ile	CTG Leu	TTA Leu 665	GAA Glu	TTT Phe	ATA Ile	CCA Pro	GAG Glu 670	ATT Ile	GCA Ala	2016
ATA Ile	CCT Pro	GTA Val 675	TTA Leu	GGT Gly	ACT Thr	TTT Phe	GCA Ala 680	CTT Leu	GTA Val	TCA Ser	TAT Tyr	ATT Ile 685	GCG Ala	AAT Asn	AAG Lys	2064
GTT Val	CTA Leu 690	ACC Thr	GTT Val	CAA Gln	ACA Thr	ATA Ile 695	GAT Asp	AAT Asn	GCT Ala	TTA Leu	AGT Ser 700	AAA Lys	AGA Arg	AAT Asn	GAA Glu	2112
AAA Lys 705	TGG Trp	GAT Asp	GAG Glu	GTC Val	TAT Tyr 710	AAA Lys	TAT Tyr	ATA Ile	GTA Val	ACA Thr 715	AAT Asn	TGG Trp	TTA Leu	GCA Ala	AAG Lys 720	2160
GTT Val	AAT Asn	ACA Thr	CAG Gln	ATT Ile 725	GAT Asp	CTA Leu	ATA Ile	AGA Arg	AAA Lys 730	AAA Lys	ATG Met	AAA Lys	GAA Glu	GCT Ala 735	TTA Leu	2208
GAA Glu	AAT Asn	CAA Gln	GCA Ala 740	GAA Glu	GCA Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT Tyr	CAG Gln 750	TAT Tyr	AAT Asn	2256
CAA Gln	TAT Tyr	ACT Thr 755	GAG Glu	GAA Glu	GAG Glu	AAA Lys	AAT Asn 760	AAT Asn	ATT Ile	AAT Asn	TTT Phe	AAT Asn 765	ATT Ile	GAT Asp	GAT Asp	2304
TTA Leu	AGT Ser 770	TCG Ser	AAA Lys	CTT Leu	AAT Asn	GAG Glu 775	TCT Ser	ATA Ile	AAT Asn	AAA Lys	GCT Ala 780	ATG Met	ATT Ile	AAT Asn	ATA Ile	2352
AAT Asn 785	AAA Lys	TTT Phe	TTG Leu	AAT Asn	CAA Gln 790	TGC Cys	TCT Ser	GTT Val	TCA Ser	TAT Tyr 795	TTA Leu	ATG Met	AAT Asn	TCT Ser	ATG Met 800	2400
ATC Ile	CCT Pro	TAT Tyr	GGT Gly	GTT Val 805	AAA Lys	CGG Arg	TTA Leu	GAA Glu	GAT Asp 810	TIT Phe	GAT Asp	GCT Ala	AGT Ser	CTT Leu 815	AAA Lys	2448
gat Asp	GCA Ala	TTA Leu	Tra Leu 820	AAG Lys	TAT Tyr	ATA Ile	TAT Tyr	GAT Asp 825	AAT Asn	AGA Arg	GGA Gly	ACT Thr	TTA Leu 830	ATT Ile	GGT Gly	2496
CAA Gln	GTA Val	GAT Asp 835	AGA Arg	TTA Leu	AAA Lys	GAT Asp	AAA Lys 840	GTT Val	AAT Asn	AAT Asn	ACA Thr	CTT Leu 845	AGT Ser	ACA Thr	GAT Asp	2544
Ile	CCT Pro 850	TTT Phe	CAG Gln	CTT Leu	TCC Ser	AAA Lys 855	TAC Tyr	GTA Val	GAT Asp	AAT Asn	CAA Gln 860	AGA Arg	TTA Leu	TTA Leu	TCT Ser	2592
	Phe				ATT Ile 870											2613

⁽²⁾ INFORMATION FOR SEQ ID NO: 8:

⁽i) SEQUENCE CHARACTERISTICS:

⁽A) LENGTH: 871 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Met Pro Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 120 Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 215 Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 265 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 280 Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 295 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu

Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu 470 Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu 490 Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro 505 Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu 570 Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Pro Tyr Ile Gly Pro Ala 630 Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys

WO 98/07864 PCT/GB97/02273

- 54 -

Val	Leu 690	Thr	Val	Gln	Thr	Ile 695	Asp	Asn	Ala	Leu	Ser 700	Lys	Arg	Asn	Glu	
Lys 705	Trp	Asp	Glu	Val	Tyr 710	Lys	Tyr	Ile	Val	Thr 715	Asn	Trp	Leu	Ala	Lys 720	
Val	Asn	Thr	Gln	11e 725	Asp	Leu	Ile	Arg	Lys 730	Lys	Met	Lys	Glu	Ala 735	Leu	
Glu	Asn	Gln	Ala 740	Glu	Ala	Thr	Lys	Ala 745	Ile	Ile	Asn	Tyr	Gln 750	Tyr	Asn	
Gln	Tyr	Thr 755	Glu	Glu	Glu	Lys ·	Asn 760	Asn	Ile	Asn	Phe	Asn 765	Ile	Asp	Asp	
Leu	Ser 770	Ser	Lys	Leu	Asn	Glu 775	Ser	Ile	Asn	Lys	Ala 780	Met	Ile	Asn	Ile	
Asn 785	Lys	Phe	Leu	Asn	Gln 790	Cys	Ser	Val	Ser	Tyr 795	Leu	Met	Asn	Ser	Met 800	
Ile	Pro	Tyr	Gly	Val 805	Lys	Arg	Leu	Glu	Asp 810	Phe	Asp	Ala	Ser	Leu 815	Lys	
Asp	Ala	Leu	Leu 820	Lys	Tyr	Ile	Tyr	Asp 825	Asn	Arg	Gly	Thr	Leu 830	Ile	Gly	
Gln	Val	Asp 835	Arg	Leu	Lys	Asp	Lys 840	Val	Asn	Asn	Thr	Leu 845	Ser	Thr	Asp	
Ile	Pro 850	Phe	Gln	Leu	`Ser	Lys 855	Tyr	Val	Asp	Asn	Gln 860	Arg	Leu	Leu	Ser	
Thr 865	Phe	Thr	Glu	Tyr	Ile 870	Lys										
(2)	INFO	RMAI	NOI	FOR	SEQ	ID N	10: 5) :								
	(i)	(E (C	QUENC A) LE B) TY C) ST D) TO	NGTI PE: RANI	i: 26 nucl EDNE	28 t eic SS:	ase acid doub	pair l	:s							
	(ii)	MOI	ECUI	E T	PE:	DNA	(ger	omic	:)							
	(ix)	(P	ATURE A) NA B) LC	ME/I			28									
	(xi)	SEC	UENC	E DE	ESCRI	PTIC	N: 5	SEQ I	D NC): 9:	:					
ATG Met 1	CAG Gln	TTC Phe	GTG Val	AAC Asn 5	AAG Lys	CAG Gln	TTC Phe	AAC	TAT Tyr 10	AAG Lys	GAC Asp	CCT Pro	GTA Val	AAC Asn 15	GGT Gly	48
			GCC Ala 20													96
			TTC Phe													144

GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG CCG GAA
-Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu
50 55 60

192

GCA Ala 65	Lys	CAG Gln	GTG Val	CCA Pro	GTT Val 70	Ser	TAC	TAC	GAT Asp	TCA Ser 75	Thr	TAT	r CTC	G AGO	ACA Thr 80	. 240
GAC Asp	AAC Asn	GAG Glu	AAG Lys	GAT Asp 85	AAC Asn	TAC	CTG Leu	AAG Lys	GGA Gly 90	Val	ACC	Lys	TTA Leu	TTO Phe 95	GAG Glu	288
CGT Arg	ATT Ile	TAT Tyr	TCC Ser 100	ACT Thr	GAC Asp	CTG Leu	GGC Gly	CGT Arg 105	Met	CTG Leu	CTG Leu	ACC	TCA Ser 110	· Ile	GTC Val	336
CGC Arg	GGA Gly	ATC Ile 115	CCA Pro	TTT Phe	TGG Trp	GGT Gly	GGC Gly 120	AGT Ser	ACC Thr	ATT Ile	GAC Asp	ACG Thr 125	Glu	TTG Leu	AAG Lys	384
GTT Val	ATT Ile 130	GAC Asp	ACT Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro 140	GAC Asp	GGT Gly	AGC Ser	TAC Tyr	432
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	CCC Pro	TCC	GCG Ala	GAC Asp	ATT Ile 160	480
ATC Ile	CAG Gln	TTT Phe	GAG Glu	TGC Cys 165	AAG Lys	AGC Ser	TTT Phe	GGC Gly	CAC His 170	GAA Glu	GTG Val	TTG Leu	AAC Asn	CTG Leu 175	ACG Thr	528
CGT Arg	AAC Asn	GGT Gly	TAC Tyr 180	GGC Gly	TCT Ser	ACT Thr	CAG Gln	TAC Tyr 185	ATT Ile	CGT Arg	TTC Phe	AGC Ser	CCA Pro 190	GAC Asp	TTC Phe	576
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
GGT Gly	GCA Ala 210	GGC	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720
CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys	816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn	864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GTG Val	912
GGT Gly 305	ACC Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	Glu	AAA Lys 320	960
TAT Tyr	CTC Leu	CTA Leu-	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	Asp	AAA Lys 335	TTA Leu	1008

	AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
	AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104
	TTT Phe	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	1152
	ACA Thr 385	ATA Ile	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
														ACT Thr			1248
	AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296
														TAC Tyr			1344
														AAT Asn			1392
														GAT Asp			1440
	AAA Lys	GGA Gly	GAA Glu	GAA Glu	ATT Ile 485	ACA Thr	TCT Ser	GAT Asp	ACT Thr	AAT Asn 490	ATA Ile	GAA Glu	GCA Ala	GCA Ala	GAA Glu 495	GAA Glu	1488
	AAT Asn	ATT Ile	AGT Ser	TTA Leu 500	GAT Asp	TTA Leu	ATA Ile	CAA Gln	CAA Gln 505	TAT Tyr	TAT Tyr	TTA Leu	ACC Thr	TTT Phe 510	AAT Asn	TTT Phe	1536
	GAT Asp	AAT Asn	GAA Glu 515	CCT Pro	GAA Glu	AAT Asn	ATT Ile	TCA Ser 520	ATA Ile	GAA Glu	AAT Asn	CTT Leu	TCA Ser 525	AGT Ser	GAC Asp	ATT Ile	1584
	ATA Ile	GGC Gly 530	CAA Gln	TTA Leu	GAA Glu	CTT Leu	ATG Met 535	CCT Pro	AAT Asn	ATA Ile	GAA Glu	AGA Arg 540	TTT Phe	CCT Pro	AAT Asn	GGA Gly	1632
	AAA Lys 545	AAG Lys	TAT Tyr	GAG Glu	TTA Leu	GAT Asp 550	AAA Lys	TAT Tyr	ACT Thr	ATG Met	TTC Phe 555	CAT His	TAT Tyr	CTT Leu	CGT Arg	GCT Ala 560	1680
	CAA Gln	GAA Glu	TTT	GAA Glu	CAT His 565	GGT Gly	AAA Lys	TCT Ser	AGG Arg	ATT Ile 570	GCT Ala	TTA Leu	ACA Thr	AAT Asn	TCT Ser 575	GTT Val	1728
	AAC Asn	GAA Glu	GCA Ala	TTA Leu 580	TTA Leu	AAT Asn	CCT Pro	AGT Ser	CGT Arg 585	GTT Val	TAT Tyr	ACA Thr	TTT Phe	TTT Phe 590	TCT Ser	TCA Ser	1776
-	GAC _Asp_	TAT Tyr	GTA Val 595	AAG Lys	AAA Lys	GTT Val	AAT Asn	AAA Lys 600	GCT Ala	ACG Thr	GAG Glu	GCA Ala	GCT -Ala- 605	ATG Met	TTT -Phe-	TTA Leu—	1824

PCT/GB97/02273 - 57 -

GGC Gly	TGG Trp 610	GTA Val	GAA Glu	CAA Gln	TTA Leu	GTA Val 615	TAT Tyr	GAT Asp	TTT Phe	ACC Thr	GAT Asp 620	GAA Glu	ACT Thr	AGC Ser	GAA Glu	1872
GTA Val 625	AGT Ser	ACT Thr	ACG Thr	GAT Asp	AAA Lys 630	ATT Ile	GCG Ala	GAT Asp	ATA Ile	ACT Thr 635	ATA Ile	ATT	ATT	CCA Pro	TAT Tyr 640	1920
ATA Ile	GGA Gly	CCT Pro	GCT Ala	TTA Leu 645	AAT Asn	ATA Ile	GGT Gly	AAT Asn	ATG Met 650	TTA Leu	TAT Tyr	AAA Lys	GAT Asp	GAT Asp 655	TTT Phe	1968
GTA Val	GGT Gly	GCT Ala	TTA Leu 660	ATA Ile	TTT Phe	TCA Ser	GGA Gly	GCT Ala 665	GTT Val	ATT Ile	CTG Leu	TTA Leu	GAA Glu 670	TTT Phe	ATA Ile	2016
CCA Pro	GAG Glu	ATT Ile 675	GCA Ala	ATA Ile	CCT Pro	GTA Val	TTA Leu 680	GGT Gly	ACT Thr	TTT Phe	GCA Ala	CTT Leu 685	GTA Val	TCA Ser	TAT Tyr	2064
ATT Ile	GCG Ala 690	AAT Asn	AAG Lys	GTT Val	CTA Leu	ACC Thr 695	GTT Val	CAA Gln	ACA Thr	ATA Ile	GAT Asp 700	AAT nsA,	GCT Ala	TTA Leu	AGT Ser	2112
			GAA Glu													2160
			AAG Lys													2208
			TTA Leu 740													2256
			AAT Asn													2304
			GAT Asp													2352
			ATA Ile													2400
			ATG Met													2448
			AAA Lys 820													2496
			GGT Gly													2544
			gat Asp													2592
	Leu	Leu	TCT Ser	Thr	Phe	Thr	Glu	Tyr	Ile	Lys	TAA					2628

__290_

(2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 876 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 260 265 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Tyr Asn 280 Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val

<u> 295</u>

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys 310 305 Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 345 Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 395 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ser Ala Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp 455 Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu 485 490 Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe 505 Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly 535 Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val 565 Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe-Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr 630 635 Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe 645

Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr 840 Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys 870

(2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2637 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE_TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..2637
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

1 5 10 15

GTT GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

96

- 61	-
------	---

														GAA Glu	CGC Arg		144
GAT Asp	ACA Thr 50	TTT Phe	ACG Thr	AAC Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	CCG Pro	CCG Pro	CCG Pro	GAA Glu		192
														AGC Ser			240
GAC Asp	AAC Asn	GAG Glu	AAG Lys	GAT Asp 85	AAC Asn	TAC Tyr	CTG Leu	AAG Lys	GGA Gly 90	GTG Val	ACC Thr	AAA Lys	TTA Leu	TTC Phe 95	GAG Glu		288
CGT Arg	ATT	TAT Tyr	TCC Ser 100	ACT Thr	GAC Asp	CTG Leu	GGC Gly	CGT Arg 105	ATG Met	CTG Leu	CTG Leu	ACC Thr	TCA Ser 110	ATC Ile	GTC Val	:	336
														TTG Leu		:	384
														AGC Ser		•	432
														GAC Asp		4	480
														CTG Leu 175			528
														GAC Asp		Ĭ	576
														CTG Leu		6	524
														CAC His		6	572
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	7	720
														GGT Gly 255		7	768
														GCG Ala		ε	316
														TAC Tyr		8	364
		Lys	Asp		Ala	Ser	Thr	Leu	Asn	Lys	Ala	Lys		ATT Ile		9	12

														GAG Glu		960
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
														GAG Glu		1056
														TTG Leu		1104
														AAT Asn		1152
														GCA Ala		1200
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	1248
														GTA Val		1296
	-													AAT Asn		1344
														AAA Lys		1392
														ACT Thr		1440
														GAA Glu 495		1488
														TTA Leu		1536
														CTT Leu		1584
AGT Ser	GAC Asp 530	ATT Ile	ATA Ile	GGC Gly	CAA Gln	TTA Leu 535	GAA Glu	CTT Leu	ATG Met	CCT Pro	AAT Asn 540	ATA Ile	GAA Glu	AGA Arg	TTT Phe	1632
														CAT His		1680
														TTA Leu 575		1728

- 63 -

AAT Asn	TCT Ser	GTT Val	AAC Asn 580	GAA Glu	GCA Ala	TTA Leu	TTA Leu	AAT Asn 585	CCT Pro	AGT Ser	CGT Arg	GTT Val	TAT Tyr 590	ACA Thr	TTT Phe	1776
TTT Phe	TCT Ser	TCA Ser 595	GAC Asp	TAT Tyr	GTA Val	AAG Lys	AAA Lys 600	GTT Val	AAT Asn	AAA Lys	GCT Ala	ACG Thr 605	GAG Glu	GCA Ala	GCT Ala	1824
ATG Met	TTT Phe 610	TTA Leu	GGC Gly	TGG Trp	GTA Val	GAA Glu 615	CAA Gln	TTA Leu	GTA Val	TAT Tyr	GAT Asp 620	TTT Phe	ACC Thr	GAT Asp	GAA Glu	1872
ACT Thr 625	AGC Ser	GAA Glu	GTA Val	AGT Ser	ACT Thr 630	ACG Thr	GAT Asp	AAA Lys	ATT Ile	GCG Ala 635	GAT Asp	ATA Ile	ACT Thr	ATA Ile	ATT Ile 640	1920
ATT Ile	CCA Pro	TAT Tyr	ATA Ile	GGA Gly 645	CCT Pro	GCT Ala	TTA Leu	AAT Asn	ATA Ile 650	GGT Gly	AAT Asn	ATG Met	TTA Leu	TAT Tyr 655	AAA Lys	1968
						TTA Leu										2016
						GCA Ala										2064
						AAG Lys 695										2112
						GAA Glu										2160
GTA Val	ACA Thr	AAT Asn	TGG Trp	TTA Leu 725	GCA Ala	AAG Lys	GTT Val	AAT naA	ACA Thr 730	CAG Gln	ATT Ile	GAT Asp	CTA Leu	ATA Ile 735	AGA Arg	2208
						TTA Leu										225.6
						TAA TaA										2304
						GAT Asp 775										2352
						ATA Ile										2400
						ATG Met										2448
						AAA Lys										2496
						GGT Gly										2544

AAT Asn	AAT Asn 850	ACA Thr	CTT Leu	AGT Ser	ACA Thr	GAT Asp 855	ATA Ile	CCT Pro	TTT Phe	CAG Gln	CTT Leu 860	TCC Ser	AAA Lys	TAC Tyr	GTA Val	2592
GAT Asp 865	AAT Asn	CAA Gln	AGA Arg	TTA Leu	TTA Leu 870	TCT Ser	ACA Thr	TTT Phe	ACT Thr	GAA Glu 875	TAT Tyr	ATT Ile	AAG Lys	TAA *		2637
(2)	INF	ORMA'	rion	FOR	SEQ	ID 1	10: :	12:								
		(1	SEQUIA) LI B) T' D) T'	ENGTI YPE :	H: 87	79 an	nino cid									
) MOI) SE(EQ I	ID NO): 1 2	2:					
Met 1	Gln	Phe	Val	Asn 5	Lys	Gln	Phe	Asn	Tyr 10	Lys	Asp	Pro	Val	Asn 15	Gly	
Val	Asp	Ile	Ala 20	Tyr	Ile	Lys	Ile	Pro 25	Asn	Ala	Gly	Gln	Met 30	Gln	Pro	
Val	Lys	Ala 35	Phe	Lys	Ile	His	Asn 40	Lys	Ile	Trp	Val	Ile 45	Pro	Glu	Arg	
Asp	Thr 50	Phe	Thr	Asn	Pro	Glu 55	Glu	Gly	Asp	Leu	Asn 60	Pro	Pro	Pro	Glu	
Ala 65	Lys	Gln	Val	Pro	Val 70	Ser	Tyr	Tyr	Asp	Ser 75	Thr	Tyr	Leu	Ser	Thr 80	
Asp	Asn	Glu	Lys	Asp 85	Asn	Tyr	Leu	Lys	Gly 90	Val	Thr	Lys	Leu	Phe 95	Glu	
Arg	Ile	Tyr	Ser 100	Thr	Asp	Leu	Gly	Arg 105	Met	Leu	Leu	Thr	Ser 110	Ile	Val	
Arg	Gly	Ile 115	Pro	Phe	Trp	Gly	Gly 120	Ser	Thr	Ile	Asp	Thr 125	Glu	Leu	Lys	
Val	Ile 130	Asp	Thr	Asn	Cys	Ile 135	Asn	Val	Ile	Gln	Pro 140	Asp	Gly	Ser	Tyr	
145		Glu			150					155					160	
Ile	Gln	Phe	Glu	Сув 165	Lys	Ser	Phe	Gly	His 170	Glu	Val	Leu	Asn	Leu 175	Thr	
Arg	Asn	Gly	Tyr 180	Gly	Ser	Thr	Gln	Tyr 185	Ile	Arg	Phe	Ser	Pro 190	Asp	Phe	
Thr	Phe	Gly 195	Phe	Glu	Glu	Ser	Leu 200	Glu	Val	Asp	Thr	Asn 205	Pro	Leu	Leu	
Gly	Ala 210	Gly	Lys	Phe	Ala	Thr 215	Asp	Pro ·	Ala	Val	Thr 220	Leu	Ala	His	Glu	
Leu 225	Ile	His	Ala	Gly	His 230	Arg	Leu	Tyr	Gly	Ile 235	Ala	Ile	Asn	Pro	Asn 240	

____Arg_Val Phe Lys_Val_Asn_Thr_Asn_Ala_Tyr_Tyr_Glu_Met_Ser_Gly_Leu-245 250 255

Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 265 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 295 Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 425 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn 470 Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 490 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 505 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr 570 Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 585 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala

Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 615 Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 630 Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 680 Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 710 Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 730 Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 840 Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys 870

(2) INFORMATION FOR SEQ ID NO: 13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2862 base pairs
 - (B) TYPE: nucleic acid

 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION:1..2862
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

	-					CAG Gln									GGT Gly		48
						AAA Lys											96
						CAT His											144
						GAA Glu 55											192
						TCA Ser											240
						TAC Tyr											288
						CTG Leu											336
						GGT Gly											384
						ATT Ile 135											432
						CTC Leu											480
						AGC Ser											528
						ACT Thr											576
						AGC Ser											624
						ACT Thr 215											672
						CGT Arg											720
						ACC Thr											768
GAA Glu	GTA Val	AGC Ser	Phe	GAG Glu	GAA Glu	CTG Leu	CGC Arg	Thr	TTC Phe	GGT Gly	GGC	CAT His	Asp	GCG Ala	AAG Lys	<u> </u>	816
			260					265					270				

		03.0		mma	220	~~		~~~		-						
Phe	Ile	Asp 275	Ser	Leu	Gln	Glu	Asn 280	Glu	Phe	Arg	Leu	TAC Tyr 285	TAC	TAC Tyr	AAC Asn	864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GTG Val	912
GGT Gly 305	ACC Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320	960
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT Tyr	TTG Leu	AAT Asn	1104
TTT Phe	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	1152
														GCA Ala		1200
														AAA Lys 415		1248
														GTA Val		1296
														AAT Asn		1344
														AAA Lys		1392
														ACT Thr		1440
														GAA Glu 495		1488
														TTA Leu		1536
														CTT Leu		1584
AGT Ser	GAC Asp 530	ATT Ile	ATA	GGC Gly	CAA Gln	TTA Leu 535	GAA Glu	CTT Leu	ATG Met	CCT Pro	AAT Asn 540	ATA Ile	GAA Glu	AGA Arg	TTT Phe	1632

		AAA Lys							16	80
		CAA Gln							17	28
		AAC Asn 580							17	76
		GAC Asp					 	 	18	24
		GGC Gly							18	72
•	 -	GTA Val						_	19	20
		ATA Ile							19	68
		GTA Val 660							20	16
		CCA Pro							20	64
		ATT Ile							21	12
		AAA Lys							21	60
		TGG Trp							22	80
		AAA Lys 740							22!	56
		TAT Tyr							23	04
	 _	AAT Asn					_		23	52
		ATG Met							24	00
		_ATG Met							244	4.8

				Asp				TAT Tyr		2496
								AAA Lys		2544
								TAC Tyr		2592
								TCT Ser		2640
								CTT Leu 895		2688
 						 		GGG Gly		2736
 								GAG Glu		2784
 								GCA Ala		2832
 	_		TCA Ser 950							2862

(2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 954 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly 10 1

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 90 85

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 105

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile. Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175 Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 215 Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 265 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 425 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 455 ... _ ... _460 =

Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe 535 Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 585 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 665 Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn 695 Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile 705 Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 805 810 815

- 73 -

AS]	o Phe	Asp	Ala 820	Ser	Leu	Lys	Asp	Ala 825		Leu	Lys	туг	: Ile 830	_	Asp	
Ası	n Arg	Gly 835	Thr	Leu	Ile	Gly	Gln 840	Val	Asp	Arg	Leu	Lys 845		Lys	. Val	
Ası	850	Thr	Leu	Ser	Thr	Asp 855	Ile	Pro	Phe	Gln	Leu 860		Lys	туг	Val	
Asp 865	Asn	Gln	Arg	Leu	Leu 870	Ser	Thr	Phe	Thr	Glu 875	Tyr	Ile	Lys	Ser	880	
Pro	Gly	Pro	Glu	Thr 885	Leu	Cys	Gly	Ala	Glu 890	Leu	Val	Asp	Ala	Leu 895		
Phe	val	Cys	Gly 900	Asp	Arg	Gly	Phe	Tyr 905	Phe	Asn	Lys	Pro	Thr 910		Tyr	
Gly	Ser	Ser 915	Ser	Arg	Arg	Ala	Pro 920	Gln	Thr	Gly	Ile	Val 925	Asp	Glu	Cys	
Cys	Phe 930	Arg	Ser	Cys	Asp	Leu 935	Arg	Arg	Leu	Glu	Met 940	Tyr	Cys	Ala	Pro	
Leu 945		Pro	Ala	Lys	Ser 950	Ala	Glu	Ala	*							
(2)	INF	ORMAI	CION	FOR	SEQ	ID N	10: 3	15:								
	(ii)	(B (C (D MOL FEA	LECUL	ENGTH PE: TRANE POLC E TY	i: 27 nucl EDNE GY:	24 b eic SS: line	ase acid doub ar	pair d ole								
				CATI	ON:1	27	24									
	(xi)	(B) LO					EO I	D NO	: 15	:					
ATG	CAG	(B SEQ TTC () LO UENC GTG	E DE	SCRI AAG	PTIO CAG	N: S	AAC	TAT	AAG	GAC	ССТ	GTA	AAC	GGT	48
ATG Met 1	CAG	(B) LO UENC GTG	E DE	SCRI AAG	PTIO CAG	N: S	AAC	TAT	AAG	GAC	CCT Pro	GTA Val	AAC Asn 15	GGT Gly	48
Met 1 GTT	CAG Gln GAC	(B SEQ TTC () LO UENC GTG Val	E DE AAC . Asn 5	SCRI AAG Lys (PTIO CAG Gln	N: S TTC Phe ATT	AAC Asn	TAT Tyr 10	AAG Lys GCC	GAC Asp	Pro	Val	Asn 15	CCG	4 8 96
Met 1 GTT Val	CAG Gln GAC Asp	SEQTTC (Phe T	UENC GTG Val GCC Ala 20	AAC ASn 5	SCRI AAG Lys ATC :	PTIO CAG Gln AAA Lys	N: S TTC Phe ATT Ile	AAC Asn CCA Pro 25	TAT Tyr 10 AAC Asn	AAG Lys GCC Ala	GAC Asp GGC Gly	Pro CAG Gln	Val ATG Met 30	Asn 15 CAG Gln	Gly CCG Pro	
GTT Val GTG Val	CAG Gln GAC Asp AAG Lys	SEQ TTC Phe Talle ATT GCT TAlle I	UENC GTG Val GCC Ala 20 FTC A	AAC (AAC (AAC (AAC (AAC (AAC (AAC (AAC	SCRI AAG Lys ATC ATC Ile I	PTIO CAG Gln AAA Lys CAT	N: S TTC Phe ATT Ile AAC ASD 40	AAC Asn CCA Pro 25 AAA Lys	TAT Tyr 10 AAC Asn ATC	AAG Lys GCC Ala TGG	GAC Asp GGC Gly GTT Val	Pro CAG Gln ATT Ile 45	Val ATG Met 30 CCG Pro	Asn 15 CAG Gln GAA Glu	Gly CCG Pro CGC Arg	96
GTT Val GTG Val GAT Asp	GAC Asp AAG (Lys ACA Thr 50 AAG (AAG (AAG (AAG (AAG (AAG (AAG (AAG	SEQ TTC (Phe ST)	UENC GTG Val GCC Ala 20 FTC Phe	AAC AAC AAAC AAAC AAAC AAAC AAAC AAAC	SCRI AAG Lys ATC Ile Ile ICCG CTT T	PTIO CAG Gln AAA Lys CAT His GAA (Slu 55	N: S TTC Phe ATT Ile AAC ASD 40 GAA	AAC Asn CCA Pro 25 AAA Lys GGA	TAT Tyr 10 AAC Asn ATC GAC GAC GASP	AAG Lys GCC Ala TGG Trp	GAC Asp GGC Gly GTT Val AAC Asn 60	CAG Gln ATT Ile 45 CCG Pro	ATG Met 30 CCG Pro	Asn 15 CAG Gln GAA Glu CCG Pro	Gly CCG Pro CGC Arg GAA Glu	96 144

								CGT Arg 105									336
CGC Arg	GGA Gly	ATC Ile 115	CCA Pro	TTT Phe	TGG Trp	GGT Gly	GGC Gly 120	AGT Ser	ACC Thr	ATT Ile	GAC Asp	ACG Thr 125	GAG Glu	TTG Leu	AAG Lys		384
GTT Val	ATT Ile 130	GAC Asp	ACT Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro 140	GAC Asp	GGT Gly	AGC Ser	TAC Tyr		432
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160		480
								GGC Gly									528
								TAC Tyr 185									576
								GAG Glu									624
								CCA Pro									672
								TAT Tyr									720
								GCC Ala									768
								ACG Thr 265									816
TTT Phe	ATC Ile	GAC Asp 275	AGC Ser	TTG Leu	CAG Gln	GAG Glu	AAC Asn 280	GAG Glu	TTC Phe	CGT Arg	CTG Leu	TAC Tyr 285	TAC Tyr	TAC Tyr	AAC Asn		864
AAG Lys	TTT Phe 290	AAA Lys	GAT Asp	ATT Ile	GCA Ala	AGT Ser 295	ACA Thr	CTG Leu	AAC Asn	AAG Lys	GCT Ala 300	AAG Lys	TCC Ser	ATT Ile	GTG Val		912
GGT Gly 305	ACC Thr	ACT Thr	GCT Ala	TCA Ser	TTA Leu 310	CAG Gln	TAT Tyr	ATG Met	AAA Lys	AAT Asn 315	GTT Val	TTT Phe	AAA Lys	GAG Glu	AAA Lys 320		960
TAT Tyr	CTC Leu	CTA Leu	TCT Ser	GAA Glu 325	GAT Asp	ACA Thr	TCT Ser	GGA Gly	AAA Lys 330	TTT Phe	TCG Ser	GTA Val	GAT Asp	AAA Lys 335	TTA Leu	;	1008
AAA Lys	TTT Phe	GAT Asp	AAG Lys 340	TTA Leu	TAC Tyr	AAA Lys	ATG Met	TTA Leu 345	ACA Thr	GAG Glu	ATT Ile	TAC Tyr	ACA Thr 350	GAG Glu	GAT Asp	:	1056
AAT Asn	TTT Phe	GTT Val 355	AAG Lys	TTT Phe	TTT Phe	AAA Lys	GTA Val 360	CTT Leu	AAC Asn	AGA Arg	AAA Lys	ACA Thr 365	TAT	TTG Leu	AAT Asn		1104

TTT Phe	GAT Asp 370	AAA Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC	1152
ACA Thr 385	ATA Ile	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	AAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
															CTA Leu	1248
															AGA Arg	1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
						GGG Gly 455										1392
AAT Asn 465	AAT Asn	TGG Trp	GAC Asp	TTG Leu	TTT Phe 470	TTT	AGT Ser	CCT Pro	TCA Ser	GAA Glu 475	GAT Asp	TAA Asn	TTT Phe	ACT Thr	AAT Asn 480	1440
						GAA Glu										1488
						TTA Leu										1536
						CCT Pro										1584
AGT Ser	GAC Asp 530	ATT Ile	ATA Ile	GGC Gly	CAA Gln	TTA Leu 535	GAA Glu	CTT Leu	ATG Met	CCT Pro	AAT Asn 540	ATA Ile	GAA Glu	AGA Arg	TTT Phe	1632
						GAG Glu										1680
CTT	Arg	GCT Ala	CAA Gln	GAA Glu 565	TTT Phe	GAA Glu	CAT His	GGT Gly	AAA Lys 570	TCT Ser	AGG Arg	ATT Ile	GCT Ala	TTA Leu 575	ACA Thr	1728
						TTA Leu	Leu	Asn		Ser	Arg					1776
						AAG Lys										1824
						GAA Glu 615										1872
						ACG . Thr									ATT Ile 640	1920

	ATT Ile	CCA Pro	TAT Tyr	ATA Ile	GGA Gly 645	CCT Pro	GCT Ala	TTA Leu	AAT Asn	ATA Ile 650	GGT Gly	AAT Asn	ATG Met	TTA Leu	TAT Tyr 655	AAA Lys	1968
	GAT Asp	GAT Asp	TTT Phe	GTA Val 660	GGT Gly	GCT Ala	TTA Leu	ATA Ile	TTT Phe 665	TCA Ser	GGA Gly	GCT Ala	GTT Val	ATT Ile 670	CTG Leu	TTA Leu	2016
	GAA Glu	TTT Phe	ATA Ile 675	CCA Pro	GAG Glu	ATT Ile	GCA Ala	ATA Ile 680	CCT Pro	GTA Val	TTA Leu	GGT Gly	ACT Thr 685	TTT Phe	GCA Ala	CTT Leu	2064
	GTA Val	TCA Ser 690	TAT Tyr	ATT Ile	GCG Ala	AAT Asn	AAG Lys 695	GTT Val	CTA Leu	ACC Thr	GTT Val	CAA Gln 700	ACA Thr	ATA Ile	GAT Asp	AAT Asn	2112
	GCT Ala 705	TTA Leu	AGT Ser	AAA Lys	AGA Arg	AAT Asn 710	GAA Glu	AAA Lys	TGG Trp	GAT Asp	GAG Glu 715	GTC Val	TAT Tyr	AAA Lys	TAT Tyr	ATA Ile 720	2160
	GTA Val	ACA Thr	AAT Asn	TGG Trp	TTA Leu 725	GCA Ala	AAG Lys	GTT Val	AAT Asn	ACA Thr 730	CAG Gln	ATT Ile	GAT Asp	CTA Leu	ATA Ile 735	AGA Arg	2208
	AAA Lys	AAA Lys	ATG Met	AAA Lys 740	GAA Glu	GCT Ala	TTA Leu	GAA Glu	AAT Asn 745	CAA Gln	GCA Ala	GAA Glu	GCA Ala	ACA Thr 750	AAG Lys	GCT Ala	2256
	ATA Ile	ATA Ile	AAC Asn 755	TAT Tyr	CAG Gln	TAT Tyr	AAT Asn	CAA Gln 760	TAT Tyr	ACT Thr	GAG Glu	GAA Glu	GAG Glu 765	AAA Lys	AAT Asn	AAT Asn	2304
	ATT Ile	AAT Asn 770	TTT Phe	AAT Asn	ATT Ile	GAT Asp	GAT Asp 775	TTA Leu	AGT Ser	TCG Ser	AAA Lys	CTT Leu 780	AAT Asn	GAG Glu	TCT Ser	ATA Ile	2352
															TCT Ser		2400
	TCA Ser	TAT Tyr	TTA Leu	ATG Met	TAA neA 208	TCT Ser	ATG Met	ATC Ile	CCT Pro	TAT Tyr 810	GGT Gly	GTT Val	AAA Lys	CGG Arg	TTA Leu 815	GAA Glu	2448
															TAT Tyr		2496
															AAA Lys		2544
															TAC Tyr		2592
															TCT Ser		2640
															TCT Ser 895		2688
-	ATT	GAT-	-ACA	CAT	TAA	-AGA-	ATT-	-AAG-	-GAT-	GAA-	TTA-	-TGA-					2724
	Ile	Asp	Thr	900	Asn	Arg	тте	пÀг	905	GIU	ren	*					

- (2) INFORMATION FOR SEQ ID NO: 16:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 908 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: protein
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Glu Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 135 Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe 185 Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 215 Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 265

Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val 290 295 300

Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn

285

Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 360 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 395 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 425 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val 455 Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 490 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr 545 550 555 Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe 585 Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala 600 Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile 635 Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys ----645----

PCT/GB97/02273

Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu 665 Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 840 Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 855 Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Arg Pro Gln Ser Lys Val Lys Arg Gln Ile Phe Ser Gly Tyr Gln Ser Asp Ile Asp Thr His Asn Arg Ile Lys Asp Glu Leu

- (2) INFORMATION FOR SEQ ID NO: 17:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3042 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (ix) FEATURE:
 - (A) NAME/KEY: CDS
 - (B) LOCATION: 1..3042
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT 48

Met Gln-Phe-Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

1 10 15

										-						
GTT Val	GAC Asp	ATT Ile	GCC Ala 20	Tyr	ATC	AAA Lys	ATT	CCA Pro 25	Asn	GCC Ala	GGC Gly	CAG Gln	ATG Met 30	CAG Gln	CCG Pro	96
GTG Val	AAG Lys	GCT Ala 35	Pue	AAG Lys	ATT	CAT His	AAC Asn 40	AAA Lys	ATC Ile	TGG Trp	GTT Val	ATT Ile 45	CCG Pro	GAA Glu	CGC Arg	144
GAT Asp	ACA Thr 50	Pne	ACG Thr	AAC Asn	CCG Pro	GAA Glu 55	GAA Glu	GGA Gly	GAC Asp	TTG Leu	AAC Asn 60	CCG Pro	CCG Pro	CCG Pro	GAA Glu	192
GCA Ala 65	AAG Lys	CAG Gln	GTG Val	CCA Pro	GTT Val 70	TCA Ser	TAC Tyr	TAC Tyr	GAT Asp	TCA Ser 75	ACC Thr	TAT Tyr	CTG Leu	AGC Ser	ACA Thr 80	240
GAC Asp	AAC Asn	GAG Glu	AAG Lys	GAT Asp 85	AAC Asn	TAC Tyr	CTG Leu	AAG Lys	GGA Gly 90	GTG Val	ACC Thr	AAA Lys	TTA Leu	TTC Phe 95	GAG Glu	288
CGT Arg	ATT Ile	TAT Tyr	TCC Ser 100	ACT Thr	GAC Asp	CTG Leu	GGC Gly	CGT Arg 105	ATG Met	CTG Leu	CTG Leu	ACC Thr	TCA Ser 110	ATC Ile	GTC Val	336
CGC Arg	GGA Gly	ATC Ile 115	CCA Pro	TTT Phe	TGG Trp	GGT Gly	GGC Gly 120	AGT Ser	ACC Thr	ATT Ile	GAC Asp	ACG Thr 125	GAG Glu	TTG Leu	AAG Lys	384
GTT Val	ATT Ile 130	GAC Asp	ACT Thr	AAC Asn	TGC Cys	ATT Ile 135	AAC Asn	GTG Val	ATC Ile	CAA Gln	CCA Pro 140	GAC Asp	GGT Gly	AGC Ser	TAC Tyr	432
AGA Arg 145	TCT Ser	GAA Glu	GAA Glu	CTT Leu	AAC Asn 150	CTC Leu	GTA Val	ATC Ile	ATC Ile	GGG Gly 155	CCC Pro	TCC Ser	GCG Ala	GAC Asp	ATT Ile 160	480
Ile	Gin	TTT	Glu	Cys 165	Lys	Ser	Phe	Gly	His 170	Glu	Val	Leu	Asn	Leu 175	Thr	528
Arg	Asn	GGT Gly	180	Gly	Ser	Thr	Gln	Tyr 185	Ile	Arg	Phe	Ser	Pro 190	Asp	Phe	576
ACG Thr	TTC Phe	GGT Gly 195	TTC Phe	GAG Glu	GAG Glu	AGC Ser	CTG Leu 200	GAG Glu	GTT Val	GAT Asp	ACC Thr	AAC Asn 205	CCG Pro	CTG Leu	TTG Leu	624
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT Tyr	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720
CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
GAA Glu	GTA Val	AGC Ser	TTC Phe 260	GAG Glu	GAA Glu	CTG Leu	CGC Arg	ACG Thr 265	TTC Phe	GGT Gly	GGC Gly	CAT His	GAT Asp 270	GCG Ala	AAG Lys	816
TTT- Phe	ATC-	GAC Asp	AGC Ser	TTG Leu	CAG- Gln	GAG- Glu	Asn	GAG- Glu	TTC Phe	CGT- Arg	CTG- Leu	TAC- Tyr	TAC- Tyr	TAC- Tyr	AAC Asn	864
		275					280					285				

						AGT Ser 295									GTG Val	912
						CAG Gln										960
						ACA Thr										1008
						AAA Lys										1056
						AAA Lys										1104
						AAG Lys 375										1152
						AAT Asn										1200
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	AAT Asn	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	1248
						TTT Phe										1296
						ACT Thr										1344
						GGG Gly 455										1392
						TTT Phe										1440
						GAA Glu										1488
						TTA Leu										1536
						CCT Pro										1584
AGT Ser	GAC Asp 530	ATT Ile	ATA Ile	GGC Gly	CAA Gln	TTA Leu 535	GAA Glu	CTT Leu	ATG Met	CCT Pro	AAT Asn 540	ATA Ile	GAA Glu	AGA Arg	TTT Phe	1632
						GAG Glu									TAT- Tyr 560	 1680

- 82 -

CTT Leu	CGT Arg	GCT Ala	CAA Gln	GAA Glu 565	TTT Phe	GAA Glu	CAT His	GGT Gly	AAA Lys 570	TCT Ser	AGG Arg	ATT Ile	GCT Ala	TTA Leu 575	ACA Thr	1728
AAT Asn	TCT Ser	GTT Val	AAC Asn 580	GAA Glụ	GCA Ala	TTA Leu	TTA Leu	AAT Asn 585	CCT Pro	AGT Ser	CGT Arg	GTT Val	TAT Tyr 590	ACA Thr	TTT Phe	1776
												ACG Thr 605				1824
												TTT Phe				1872
												ATA Ile				1920
												ATG Met				1968
												GTT Val				2016
												ACT Thr 685				2064
												ACA Thr				2112
												TAT Tyr				2160
GTA Val	ACA Thr	AAT Asn	TGG Trp	TTA Leu 725	GCA Ala	AAG Lys	GTT Val	AAT Asn	ACA Thr 730	CAG Gln	ATT Ile	GAT Asp	CTA Leu	ATA Ile 735	AGA Arg	2208
AAA Lys	Lys	Met	Lys	Glu	Ala	Leu	Glu	Asn	Gln	Ala	Glu	GCA Ala	Thr	Lys	GCT Ala	2256
ATA Ile	ATA Ile	AAC Asn 755	TAT Tyr	CAG Gln	TAT Tyr	AAT Asn	CAA Gln 760	TAT Tyr	ACT Thr	GAG Glu	GAA Glu	GAG Glu 765	AAA Lys	AAT Asn	AAT Asn	2304
												AAT Asn				2352
AAT Asn 785	AAA Lys	GCT Ala	ATG Met	ATT Ile	AAT Asn 790	ATA Ile	AAT Asn	AAA Lys	TTT Phe	TTG Leu 795	AAT Asn	CAA Gln	TGC Cys	TCT Ser	GTT Val 800	2400
TCA Ser	TAT Tyr	TTA Leu	ATG Met	AAT Asn 805	TCT Ser	ATG Met	ATC Ile	CCT Pro	TAT Tyr 810	GGT Gly	GTT Val	AAA Lys	CGG Arg	TTA Leu 815	GAA Glu	2448
GAT Asp	TTT Phe	GAT Asp	GCT Ala 820	AGT Ser	CTT Leu	AAA Lys	GAT Asp	GCA Ala 825	TTA Leu	TTA Leu	AAG Lys	TAT Tyr	ATA Ile 830	TAT	GAT Asp	2496

AAT Asn	AGA Arg	GGA Gly 835	ACT Thr	TTA Leu	ATT Ile	GGT Gly	CAA Gln 840	GTA Val	GAT Asp	AGA Arg	TTA Leu	AAA Lys 845	GAT Asp	AAA Lys	GTT Val	2544
AAT Asn	AAT Asn 850	ACA Thr	CTT Leu	AGT Ser	ACA Thr	GAT Asp 855	ATA Ile	CCT Pro	TTT Phe	CAG Gln	CTT Leu 860	TCC Ser	AAA Lys	TAC Tyr	GTA Val	2592
GAT Asp 865	AAT Asn	CAA Gln	AGA Arg	TTA Leu	TTA Leu 870	TCT Ser	ACA Thr	TTT Phe	ACT Thr	GAA Glu 875	TAT Tyr	ATT Ile	AAG Lys	TCA Ser	GGC Gly 880	2640
CTG Leu	AAT Asn	TCC Ser	CCG Pro	GGT Gly 885	GCA Ala	GCT Ala	CAT His	TAT Tyr	GCG Ala 890	CAA Gln	CAC His	GAT Asp	GAA Glu	GCC Ala 895	GTA Val	2688
GAC Asp	AAC Asn	AAA Lys	TTC Phe 900	AAC Asn	AAA Lys	GAA Glu	CAA Gln	CAA Gln 905	AAC Asn	GCG Ala	TTC Phe	TAT Tyr	GAG Glu 910	ATC Ile	TTA Leu	2736
CAT His	TTA Leu	CCT Pro 915	AAC Asn	TTA Leu	AAC Asn	GAA Glu	GAA Glu 920	CAA Gln	CGA Arg	AAC Asn	GCC Ala	TTC Phe 925	ATC Ile	CAA Gln	AGT Ser	2784
TTA Leu	AAA Lys 930	GAT Asp	GAC Asp	CCA Pro	AGC Ser	CAA Gln 935	AGC Ser	GCT Ala	AAC Asn	CTT Leu	TTA Leu 940	GCA Ala	GAA Glu	GCT Ala	AAA Lys	2832
AAG Lys 945	CTA Leu	AAT Asn	GAT Asp	GCT Ala	CAG Gln 950	GCG Ala	CCG Pro	AAA Lys	GTA Val	GAC Asp 955	AAC Asn	AAA Lys	TTC Phe	AAC Asn	AAA Lys 960	2880
GAA Glu	CAA Gln	CAA Gln	AAC Asn	GCG Ala 965	TTC Phe	TAT Tyr	GAG Glu	ATC Ile	TTA Leu 970	CAT His	TTA Leu	CCT Pro	AAC Asn	TTA Leu 975	AAC Asn	2928
GAA Glu	GAA Glu	CAA Gln	CGA Arg 980	AAC Asn	GCC Ala	TTC Phe	ATC Ile	CAA Gln 985	AGT Ser	TTA Leu	AAA Lys	GAT Asp	GAC Asp 990	CCA Pro	AGC Ser	2976
CAA Gln	AGC Ser	GCT Ala 995	AAC Asn	CTT Leu	TTA Leu	Ala	GAA Glu 1000	Ala	AAA Lys	AAG Lys	CTA Leu	AAT Asn 1005	Asp	GCT Ala	CAG Gln	3024
	CCG Pro 1010	Lys			TAG											3042

(2) INFORMATION FOR SEQ ID NO: 18:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1014 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 40

Asp	Thr 50	Phe	Thr	Asn	Pro	Glu 55	Glu	Gly	Asp	Leu	Asn 60	Pro	Pro	Pro	Glu
Ala 65	ŗÀs	Gln	Val	Pro	Val 70	Ser	Tyr	Tyr	Asp	Ser 75	Thr	Tyr	Leu	Ser	Thr 80
Asp	Asn	Glu	Lys	Asp 85	Asn	Tyr	Leu	Lys	Gly 90	Val	Thr	Lys	Leu	Phe 95	Glu
Arg	Ile	Tyr	Ser 100	Thr	Asp	Leu	Gly	Arg 105	Met	Leu	Leu	Thr	Ser 110	Ile	Val
Arg	Gly	Ile 115	Pro	Phe	Trp	Gly	Gly 120	Ser	Thr	Ile	Asp	Thr 125	Glu	Leu	Lys
Val	Ile 130	Asp	Thr	Asn	Cys	Ile 135	Asn	Val	Ile	Gln	Pro 140	Asp	Gly	Ser	Tyr
Arg 145	Ser	Glu	Glu	Leu	Asn 150	Leu	Val	Ile	Ile	Gly 155	Pro	Ser	Ala	Asp	11e 160
Ile	Gln	Phe	Glu	Cys 165	Lys	Ser	Phe	Gly	His 170	Glu	Val	Leu	Asn	Leu 175	Thr
Arg	Asn	Gly	Tyr 180	Gly	Ser	Thr	Gln	Tyr 185	Ile	Arg	Phe	Ser	Pro 190	Asp	Phe
Thr	Phe	Gly 195	Phe	Glu	Glu	Ser	Leu 200	Glu	Val	Asp	Thr	Asn 205	Pro	Leu	Leu
Gly	Ala 210	Gly	Lys	Phe	Ala	Thr 215	Asp	Pro	Ala	Val	Thr 220	Leu	Ala	His	Glu
Leu 225	Ile	His	Ala	Gly	His 230	Arg	Leu	Tyr	Gly	Ile 235	Ala	Ile	Asn	Pro	Asn 240
				245					250					Gly 255	
Glu	Val	Ser	Phe 260	Glu	Glu	Leu	Arg	Thr 265	Phe	Gly	Gly	His	Asp 270	Ala	Lys
Phe	Ile	Asp 275	Ser	Leu	Gln	Glu	Asn 280	Glu	Phe	Arg	Leu	Tyr 285	Tyr	Tyr	Asn
Lys	Phe 290	Lys	Asp	Ile	Ala	Ser 295	Thr	Leu	Asn	Lys	Ala 300	Lys	Ser	Ile	Val
305					310					315			_	Glu	320
Tyr	Leu	Leu	Ser	Glu 325	Ąsp	Thr	Ser	Gly	Lys 330	Phe	Ser	Val	Asp	Lys 335	Leu
Lys	Phe	Asp	Lys 340	Leu	Tyr	Lys	Met	Leu 345	Thr	Glu	Ile	Tyr	Thr 350	Glu	qaA
		355					360					365		Leu	
	370					375					380			Asn	
Thr	Ile	Tyr	Asp	Gly	Phe	Asn	Leu	Arg	Asn	Thr	Asn	Leu	Ala	Ala	Asn
385					390					395					400
202															

Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg 420 Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ile Glu Gly Arg Cys Asp Gly Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe Ser Pro Ser Glu Asp Asn Phe Thr Asn Asp Leu Asn Lys Gly Glu Glu Ile Thr Ser Asp Thr Asn Ile Glu Ala 490 Ala Glu Glu Asn Ile Ser Leu Asp Leu Ile Gln Gln Tyr Tyr Leu Thr 505 Phe Asn Phe Asp Asn Glu Pro Glu Asn Ile Ser Ile Glu Asn Leu Ser Ser Asp Ile Ile Gly Gln Leu Glu Leu Met Pro Asn Ile Glu Arg Phe Pro Asn Gly Lys Lys Tyr Glu Leu Asp Lys Tyr Thr Met Phe His Tyr Leu Arg Ala Gln Glu Phe Glu His Gly Lys Ser Arg Ile Ala Leu Thr Asn Ser Val Asn Glu Ala Leu Leu Asn Pro Ser Arg Val Tyr Thr Phe Phe Ser Ser Asp Tyr Val Lys Lys Val Asn Lys Ala Thr Glu Ala Ala Met Phe Leu Gly Trp Val Glu Gln Leu Val Tyr Asp Phe Thr Asp Glu 615 · Thr Ser Glu Val Ser Thr Thr Asp Lys Ile Ala Asp Ile Thr Ile Ile Ile Pro Tyr Ile Gly Pro Ala Leu Asn Ile Gly Asn Met Leu Tyr Lys Asp Asp Phe Val Gly Ala Leu Ile Phe Ser Gly Ala Val Ile Leu Leu Glu Phe Ile Pro Glu Ile Ala Ile Pro Val Leu Gly Thr Phe Ala Leu 680 Val Ser Tyr Ile Ala Asn Lys Val Leu Thr Val Gln Thr Ile Asp Asn Ala Leu Ser Lys Arg Asn Glu Lys Trp Asp Glu Val Tyr Lys Tyr Ile Val Thr Asn Trp Leu Ala Lys Val Asn Thr Gln Ile Asp Leu Ile Arg 730 Lys Lys Met Lys Glu Ala Leu Glu Asn Gln Ala Glu Ala Thr Lys Ala 745 750 750

Ile Ile Asn Tyr Gln Tyr Asn Gln Tyr Thr Glu Glu Glu Lys Asn Asn 760 Ile Asn Phe Asn Ile Asp Asp Leu Ser Ser Lys Leu Asn Glu Ser Ile Asn Lys Ala Met Ile Asn Ile Asn Lys Phe Leu Asn Gln Cys Ser Val Ser Tyr Leu Met Asn Ser Met Ile Pro Tyr Gly Val Lys Arg Leu Glu 810 Asp Phe Asp Ala Ser Leu Lys Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly Gln Val Asp Arg Leu Lys Asp Lys Val 840 Asn Asn Thr Leu Ser Thr Asp Ile Pro Phe Gln Leu Ser Lys Tyr Val 855 Asp Asn Gln Arg Leu Leu Ser Thr Phe Thr Glu Tyr Ile Lys Ser Gly Leu Asn Ser Pro Gly Ala Ala His Tyr Ala Gln His Asp Glu Ala Val Asp Asn Lys Phe Asn Lys Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu His Leu Pro Asn Leu Asn Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser 920 Leu Lys Asp Asp Pro Ser Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys Lys Leu Asn Asp Ala Gln Ala Pro Lys Val Asp Asn Lys Phe Asn Lys 950 Glu Gln Gln Asn Ala Phe Tyr Glu Ile Leu His Leu Pro Asn Leu Asn 970 Glu Glu Gln Arg Asn Ala Phe Ile Gln Ser Leu Lys Asp Asp Pro Ser 985 Gln Ser Ala Asn Leu Leu Ala Glu Ala Lys Lys Leu Asn Asp Ala Gln 1000 1005 Ala Pro Lys Val Asp *

- (2) INFORMATION FOR SEQ ID NO: 19:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 3509 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (ix) FEATURE:

1010

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..3509
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

	ATG Met 1	Pro	GTI Val	ACA Thr	ATA Ile	AAI Asn	TAA C	TTI Phe	AAT Asr	TAT Tyr 10	Asr	GAT Asp	CC1	T ATT	GA' As ₁	r AAT p Asn 5	48
	AAT Asn	AAT Asn	ATT	ATT E Ile 20	Met	ATC Met	GAG Glu	CCT Pro	CCF Pro 25	Phe	GCG Ala	AGA Arg	GGT Gly	ACC Thi	Gly	G AGA / Arg	96
	TAT Tyr	TAT	AAA Lys 35	Ala	TTT Phe	'AAA Lys	ATC Ile	ACA Thr 40	Asp	CGT Arg	ATT Ile	TGG	ATA Ile 45	: Ile	CCC Pro	GAA Glu	144
	AGA Arg	TAT Tyr 50	Thr	TTT Phe	GGA Gly	TAT Tyr	AAA Lys 55	CCT Pro	GAG Glu	GAT Asp	TTT Phe	AAT Asn 60	Lys	AGI Ser	TCC Ser	GGT Gly	192
	ATT Ile 65	TTT Phe	AAT Asn	AGA Arg	GAT Asp	GTT Val 70	TGT Cys	GAA Glu	TAT	TAT Tyr	GAT Asp 75	Pro	GAT Asp	TAC Tyr	Leu	AAT Asn 80	240
	ACT Thr	AAT Asn	GAT Asp	AAA Lys	AAG Lys 85	Asn	ATA Ile	TTT Phe	TTA Leu	CAA Gln 90	ACA Thr	ATG Met	ATC Ile	AAG Lys	TTA Leu 95	TTT	288
	AAT Asn	AGA Arg	ATC Ile	AAA Lys 100	TCA Ser	AAA Lys	CCA Pro	TTG Leu	GGT Gly 105	GAA Glu	AAG Lys	TTA Leu	TTA Leu	GAG Glu 110	ATG Met	ATT	336
	ATA Ile	AAT Asn	GGT Gly 115	ATA Ile	CCT Pro	TAT Tyr	CTT Leu	GGA Gly 120	GAT Asp	AGA Arg	CGT	GTT Val	CCA Pro 125	CTC Leu	GAA Glu	GAG Glu	384
	Phe	Asn 130	Thr	Asn	Ile	Ala	AGT Ser 135	Val	Thr	Val	Asn	Lys 140	Leu	Ile	Ser	Asn	432
	CCA Pro 145	GGA Gly	GAA Glu	GTG Val	GAG Glu	CGA Arg 150	AAA Lys	AAA Lys	GGT Gly	ATT Ile	TTC Phe 155	GCA Ala	AAT Asn	TTA Leu	ATA Ile	ATA Ile 160	480
	TTT Phe	GGA Gly	CCT Pro	GGG Gly	CCA Pro 165	GTT Val	TTA Leu	AAT Asn	GAA Glu	AAT Asn 170	GAG Glu	ACT Thr	ATA Ile	GAT Asp	ATA Ile 175	GGT Gly	528
	ATA Ile	CAA Gln	AAT Asn	CAT His 180	TTT Phe	GCA Ala	TCA Ser	AGG Arg	GAA Glu 185	GGC Gly	TTC Phe	GGG Gly	GGT Gly	ATA Ile 190	ATG Met	CAA Gln	576
	ATG Met	AAG Lys	TTT Phe 195	TGC Cys	CCA Pro	GAA Glu	TAT Tyr	GTA Val 200	AGC Ser	GTA Val	TTT Phe	AAT Asn	AAT Asn 205	GTT Val	CAA Gln	GAA Glu	624
	Asn	AAA Lys 210	GGC Gly	GCA Ala	AGT Ser	ATA Ile	TTT Phe 215	AAT Asn	AGA Arg	CGT Arg	GGA Gly	TAT Tyr 220	TTT Phe	TCA Ser	GAT Asp	CCA Pro	672
	GCC Ala 225	TTG Leu	ATA Ile	TTA Leu	ATG Met	CAT His 230	GAA Glu	CTT Leu	ATA Ile	CAT His	GTT Val 235	TTA Leu	CAT His	GGA Gly	TTA Leu	TAT Tyr 240	720
	GGC	ATT Ile	AAA Lys	GTA Val	GAT Asp 245	GAT Asp	TTA Leu	CCA Pro	ATT Ile	GTA Val 250	CCA Pro	AAT Asn	GAA Glu	AAA Lys	AAA Lys 255	TTT Phe	768
-	TTT .	ATG Met	CAA Gln	TCT Ser 260	ACA Thr	GAT Asp	GCT Ala	Ile	CAG_ Gln 265	GCA_ Ala	GAA Glu	GAA Glu	CTA- Leu	TAT Tyr 270	ACA- Thr	TTT Phe	

GGA Gly	GGA Gly	CAA Gln 275	GAT Asp	CCC Pro	AGC Ser	ATC Ile	ATA Ile 280	ACT Thr	CCT Pro	TCT Ser	ACG Thr	GAT Asp 285	AAA Lys	AGT Ser	ATC Ile	864
TAT Tyr	GAT Asp 290	AAA Lys	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT Phe	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	GAT Asp	AGA Arg	CTT Leu	AAC Asn	912
AAG Lys	GTT Val	TTA Leu	GTT Val	TGC Cys	ATA Ile 310	TCA Ser	GAT Asp	CCT Pro	AAC Asn	ATT Ile 315	AAT Asn	ATT Ile	AAT Asn	ATA Ile	TAT Tyr 320	960
		AAA Lys														1008
		AGT Ser														1056
		GGT Gly 355														1104
		GCT Ala														1152
		TTA Leu														1200
		AAA Lys														1248
		CAA Gln														1296
		CAA Gln 435														1344
GTT Val	GAT Asp 450	TAA Asn	GAA Glu	GAT Asp	TTG Leu	TTC Phe 455	TTT Phe	ATA Ile	GCT Ala	GAT Asp	AAA Lys 460	AAT Asn	AGT Ser	TTT Phe	TCA Ser	1392
GAT Asp 465	GAT Asp	TTA Leu	TCT Ser	AAA Lys	AAC Asn 470	GAA Glu	AGA Arg	ATA Ile	GAA Glu	TAT Tyr 475	AAT Asn	ACA Thr	CAG Gln	AGT Ser	AAT Asn 480	1440
TAT Tyr	ATA Ile	GAA Glu	AAT Asn	GAC Asp 485	TTC Phe	CCT	ATA Ile	TAA naA	GAA Glu 490	TTA Leu	ATT	TTA Leu	GAT Asp	ACT Thr 495	GAT Asp	1488
TTA Leu	ATA Ile	AGT Ser	AAA Lys 500	ATA Ile	GAA Glu	TTA Leu	CCA Pro	AGT Ser 505	GAA Glu	AAT Asn	ACA Thr	GAA Glu	TCA Ser 510	CTT	ACT Thr	1536
GAT Asp	TTT Phe	AAT Asn 515	GTA Val	GAT Asp	GTT Val	CCA	GTA Val 520	Tyr	GAA Glu	AAA Lys	CAA Gln	CCC Pro 525	GCT Ala	ATA Ile	AAA Lys	1584
AAA Lys	ATT Ile 530	TTT Phe	ACA Thr	GAT Asp	GAA Glu	AAT Asn 535	ACC Thr	ATC Ile	TTT Phe	CAA Gln	TAT Tyr 540	TTA Leu	TAC	TCT Ser	CAG Gln	1632

- 89 -

ACA Thr 545	Phe	Pro	CTA Leu	GAT Asp	ATA Ile 550	Arg	GAT Asp	ATA	AGT Ser	TTA Leu 555	Thr	TCI Ser	TCA Ser	TTT Phe	GAT Asp 560	1680
GA1 Asp	GCA Ala	TTA Leu	TTA Leu	TTT Phe 565	Ser	AAC Asn	AAA Lys	GTT Val	TAT Tyr 570	Ser	TTT Phe	TTT Phe	TCT	Met 575	GAT Asp	1728
TAT Tyr	ATT Ile	AAA Lys	ACT Thr 580	Ala	AAT Asn	AAA Lys	GTG Val	GTA Val 585	Glu	GCA Ala	GGA Gly	TTA Leu	TTT Phe 590	Ala	GGT Gly	1776
TGG Trp	GTG Val	AAA Lys 595	CAG Gln	ATA Ile	GTA Val	AAT Asn	GAT Asp 600	TTT Phe	GTA Val	ATC	GAA Glu	GCT Ala 605	AAT Asn	AAA Lys	AGC Ser	1824
AAT Asn	ACT Thr 610	ATG Met	GAT Asp	AAA Lys	ATT	GCA Ala 615	GAT Asp	ATA Ile	TCT Ser	CTA Leu	ATT Ile 620	Val	CCT Pro	TAT	ATA Ile	1872
GGA Gly 625	Leu	GCT Ala	TTA Leu	AAT Asn	GTA Val 630	GGA Gly	AAT Asn	GAA Glu	ACA Thr	GCT Ala 635	AAA Lys	GGA Gly	AAT Asn	TTT Phe	GAA Glu 640	1920
AAT Asn	GCT Ala	TTT Phe	GAG Glu	ATT Ile 645	GCA Ala	GGA Gly	GCC Ala	AGT Ser	ATT Ile 650	CTA Leu	CTA Leu	GAA Glu	TTT Phe	ATA Ile 655	CCA Pro	1968
GAA Glu	CTT Leu	TTA Leu	ATA Ile 660	CCT Pro	GTA Val	GTT Val	GGA Gly	GCC Ala 665	TTT Phe	TTA Leu	TTA Leu	GAA Glu	TCA Ser 670	TAT Tyr	ATT Ile	2016
GAC Asp	AAT Asn	AAA Lys 675	AAT Asn	AAA Lys	ATT Ile	ATT Ile	AAA Lys 680	ACA Thr	ATA Ile	GAT Asp	AAT Asn	GCT Ala 685	TTA Leu	ACT Thr	AAA Lys	2064
AGA Arg	AAT Asn 690	GAA Glu	AAA Lys	TGG Trp	AGT Ser	GAT Asp 695	ATG Met	TAC Tyr	GGA Gly	TTA Leu	ATA Ile 700	GTA Val	GCG Ala	CAA Gln	TGG Trp	2112
CTC Leu 705	TCA Ser	ACA Thr	GTT Val	AAT Asn	ACT Thr 710	CAA Gln	TTT Phe	TAT Tyr	ACA Thr	ATA Ile 715	AAA Lys	GAG Glu	GGA Gly	ATG Met	TAT Tyr 720	2160
AAG Lys	GCT Ala	TTA Leu	AAT Asn	TAT Tyr 725	CAA Gln	GCA Ala	CAA Gln	GCA Ala	TTG Leu 730	GAA Glu	GAA Glu	ATA Ile	ATA Ile	AAA Lys 735	TAC Tyr	2208
Arg	Tyr	Asn	ATA Ile 740	Tyr	Ser	Glu	Lys	Glu 745	Lys	Ser	Asn	Ile	Asn 750	Ile	Asp	2256
TTT Phe	TAA NeA	GAT Asp 755	ATA Ile	AAT Asn	TCT Ser	AAA Lys	CTT Leu 760	AAT Asn	GAG Glu	GGT Gly	ATT Ile	AAC Asn 765	CAA Gln	GCT Ala	ATA Ile	2304
GAT Asp	AAT Asn 770	ATA Ile	AAT Asn	AAT Asn	TTT Phe	ATA Ile 775	AAT Asn	GGA Gly	TGT Cys	TCT Ser	GTA Val 780	TCA Ser	TAT Tyr	TTA Leu	ATG Met	2352
AAA Lys 785	AAA Lys	ATG Met	ATT Ile	CCA Pro	TTA Leu 790	GCT Ala	GTA Val	GAA Glu	AAA Lys	TTA Leu 795	CTA Leu	GAC Asp	TTT Phe	GAT Asp	AAT Asn 800	2400
ACT- Thr	CTC Leu	AAA- Lys	-AAA- Lys	AAT- Asn 805	-TTG- Leu	TTA- Leu	AAT Asn	TAT- Tyr	ATA Ile 810	GAT Asp	GAA Glu	AAT Asn	AAA Lys	TTA Leu 815	TAT Tyr	2448

- 90 -

TTG Leu	ATT Ile	GGA Gly	AGT Ser 820	GCA Ala	ĠAA Glu	TAT Tyr	GAA Glu	AAA Lys 825	TCA Ser	AAA Lys	GTA Val	AAT Asn	AAA Lys 830	TAC Tyr	TTG Leu	2496
AAA Lys	ACC Thr	ATT Ile 835	ATG Met	CCG Pro	TTT Phe	GAT Asp	CTT Leu 840	TCA Ser	ATA Ile	TAT Tyr	ACC Thr	AAT Asn 845	GAT Asp	ACA Thr	ATA Ile	2544
CTA Leu	ATA Ile 850	GAA Glu	ATG Met	TTT Phe	AAT Asn	AAA Lys 855	TAT Tyr	AAT Asn	AGC Ser	GAA Glu	ATT Ile 860	TTA Leu	AAT Asn	AAT Asn	ATT Ile	2592
ATC Ile 865	TTA Leu	AAT Asn	TTA Leu	AGA Arg	TAT Tyr 870	AAG Lys	GAT Asp	AAT Asn	AAT Asn	TTA Leu 875	ATA Ile	GAT Asp	TTA Leu	TCA Ser	GGA Gly 880	2640
TAT Tyr	GGG Gly	GCA Ala	AAG Lys	GTA Val 885	GAG Glu	GTA Val	TAT Tyr	GAT Asp	GGA Gly 890	GTC Val	GAG Glu	CTT Leu	AAT Asn	GAT Asp 895	AAA Lys	2688
AAT Asn	CAA Gln	TTT Phe	AAA Lys 900	TTA Leu	ACT Thr	AGT Ser	TCA Ser	GCA Ala 905	AAT Asn	AGT Ser	AAG Lys	ATT Ile	AGA Arg 910	GTG Val	ACT Thr	2736
CAA Gln	AAT Asn	CAG Gln 915	AAT Asn	ATC Ile	ATA Ile	TTT Phe	AAT Asn 920	AGT Ser	GTG Val	TTC Phe	CTT Léu	GAT Asp 925	TTT Phe	AGC Ser	GTT Val	2784
AGC Ser	TTT Phe 930	TGG Trp	ATA Ile	AGA Arg	ATA Ile	CCT Pro 935	AAA Lys	TAT Tyr	AAG Lys	AAT Asn	GAT Asp 940	GGT Gly	ATA Ile	CAA Gln	AAT Asn	2832
	ATT Ile															2880
	TGG Trp															2928
	ATA Ile										-		-			2976
	GAT Asp							Arg					Thr			3024
	AAT Asn 1010	Leu					Ile					Lys				3072
	ACA Thr					Ile					Ala					3120
ATA Ile	TTT Phe	AAA Lys	TTA Leu	GAT Asp 1045	Gly	GAT Asp	ATA Ile	GAT Asp	AGA Arg 1050	Thr	CAA Gln	TTT Phe	ATT Ile	TGG Trp 1059	Met	3168
	TAT Tyr			Ile					Leu					Ile		3216
GAA Glu	AGA Arg	TAT Tyr 107	Lys	ATT Ile	CAA Gln	TCA Ser	TAT Tyr 1080	Ser	GAA Glu	TAT Tyr	TTA Leu	AAA Lys 108	Asp	TTT Phe	TGG Trp	3264

GGA AA Gly As 10	T CCT in Pro	TTA Leu	ATG Met	TAC Tyr	AAT Asn 1099	Lys	GAA Glu	TAT Tyr	TAT Tyr	ATG Met 110	Phe	AAT Asn	GCG Ala	GGG Gly	•	3312
AAT AA Asn Ly 1105	A AAT 's Asn	TCA Ser	TAT Tyr	ATT Ile 1110	Lys	CTA Leu	AAG Lys	AAA Lys	GAT Asp	Ser	CCT Pro	GTA Val	GGT Gly	GAA Glu 1120		3360
ATT TT Ile Le	A ACA u Thr	CGT Arg	AGC Ser 1125	Lys	TAT Tyr	AAT Asn	CAA Gln	AAT Asn 1130	Ser	AAA Lys	TAT Tyr	ATA Ile	AAT Asn 1135	Tyr		3408
AGA GA Arg As	T TTA	TAT Tyr 1140	Ile	GGA Gly	GAA Glu	AAA Lys	TTT Phe 1145	Ile	ATA Ile	AGA Arg	AGA Arg	AAG Lys 1150	Ser	AAT Asn		3456
TCT CA Ser Gl	A TCT n Ser 1155	Ile	AAT Asn	GAT Asp	GAT Asp	ATA Ile 1160	Val	AGA Arg	AAA Lys	GAA Glu	GAT Asp 1165	Tyr	ATA Ile	TAT Tyr		3504
CTA G Leu	A											•				3509

(2) INFORMATION FOR SEQ ID NO: 20:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1169 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15

Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg 20 25 30

Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu 35 40 45

Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 60

Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80

Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe 85 90 95

Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile 100 105 110

Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu 115 120 125

Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn 130 135 140

Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile
145 150 155 160

Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly
165 170 175

Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln 185 Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 230 235 Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe 250 Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe 265 Gly Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 295 Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr 315 Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 330 Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys 360 Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys 375 Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile 390 395 Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp 440 Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 465 Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 490 Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys

Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp 550 Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu 630 Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp 695 Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr 730 Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr 805 Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser Glu Ile Leu Asn Asn Ile Ile Leu Asn Leu Arg Tyr Lys Asp Asn Asn Leu Ile Asp Leu Ser Gly 865 _____870 . ____875 --- 880

- 94 -

Tyr Gly Ala Lys Val Glu Val Tyr Asp Gly Val Glu Leu Asn Asp Lys 885 890 895

Asn Gln Phe Lys Leu Thr Ser Ser Ala Asn Ser Lys Ile Arg Val Thr 900 905 910

Gln Asn Gln Asn Ile Ile Phe Asn Ser Val Phe Leu Asp Phe Ser Val 915 920 925

Ser Phe Trp Ile Arg Ile Pro Lys Tyr Lys Asn Asp Gly Ile Gln Asn 930 935 940

Tyr Ile His Asn Glu Tyr Thr Ile Ile Asn Cys Met Lys Asn Asn Ser 945 950 955 960

Gly Trp Lys Ile Ser Ile Arg Gly Asn Arg Ile Ile Trp Thr Leu Ile 965 970 975

Asp Ile Asn Gly Lys Thr Lys Ser Val Phe Phe Glu Tyr Asn Ile Arg 980 985 990

Glu Asp Ile Ser Glu Tyr Ile Asn Arg Trp Phe Phe Val Thr Ile Thr 995 1000 1005

Asn Asn Leu Asn Asn Ala Lys Ile Tyr Ile Asn Gly Lys Leu Glu Ser 1010 1015 1020

Asn Thr Asp Ile Lys Asp Ile Arg Glu Val Ile Ala Asn Gly Glu Ile 1025 1030 1035 1040

Ile Phe Lys Leu Asp Gly Asp Ile Asp Arg Thr Gln Phe Ile Trp Met 1045 1050 1055

Lys Tyr Phe Ser Ile Phe Asn Thr Glu Leu Ser Gln Ser Asn Ile Glu 1060 1065 1070

Glu Arg Tyr Lys Ile Gln Ser Tyr Ser Glu Tyr Leu Lys Asp Phe Trp 1075 1080 1085

Gly Asn Pro Leu Met Tyr Asn Lys Glu Tyr Tyr Met Phe Asn Ala Gly 1090 1095 1100

Asn Lys Asn Ser Tyr Ile Lys Leu Lys Lys Asp Ser Pro Val Gly Glu 1105 1110 1115 1120

Ile Leu Thr Arg Ser Lys Tyr Asn Gln Asn Ser Lys Tyr Ile Asn Tyr 1125 1130 1135

Arg Asp Leu Tyr Ile Gly Glu Lys Phe Ile Ile Arg Arg Lys Ser Asn 1140 1145 1150

Ser Gln Ser Ile Asn Asp Asp Ile Val Arg Lys Glu Asp Tyr Ile Tyr 1155 1160 1165

Leu

(2) INFORMATION FOR SEQ ID NO: 21:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2574 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

WO 98/07864 PCT/GB97/02273

- 95 -

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:1..2574

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

ATG Met 1	CCA Pro	GTT Val	ACA Thr	ATA Ile 5	AAT Asn	AAT Asn	TTT Phe	AAT Asn	TAT Tyr 10	AAT Asn	GAT Asp	CCT Pro	ATT	GAT Asp 15	AAT Asn	4	1 B
AAT Asn	AAT Asn	ATT	ATT Ile 20	ATG Met	ATG Met	GAG Glu	CCT Pro	CCA Pro 25	TTT Phe	GCG Ala	AGA Arg	GGT Gly	ACG Thr 30	GGG Gly	AGA Arg	S	96
TAT Tyr	TAT Tyr	AAA Lys 35	GCT Ala	TTT	AAA Lys	ATC Ile	ACA Thr 40	GAT Asp	CGT Arg	ATT Ile	TGG Trp	ATA Ile 45	ATA Ile	CCG Pro	GAA Glu	14	4
AGA Arg	TAT Tyr 50	ACT Thr	TTT Phe	GGA Gly	TAT Tyr	AAA Lys 55	CCT Pro	GAG Glu	GAT Asp	TTT Phe	AAT Asn 60	AAA Lys	AGT Ser	TCC Ser	GGT Gly	19	2
ATT Ile 65	TTT Phe	AAT Asn	AGA Arg	GAT Asp	GTT Val 70	TGT Cys	GAA Glu	TAT Tyr	TAT Tyr	GAT Asp 75	CCA Pro	GAT Asp	TAC Tyr	TTA Leu	AAT Asn 80	24	0
ACT Thr	AAT Asn	GAT Asp	AAA Lys	AAG Lys 85	AAT Asn	ATA Ile	TTT Phe	TTA Leu	CAA Gln 90	ACA Thr	ATG Met	ATC Ile	AAG Lys	TTA Leu 95	TTT Phe	28	8
AAT Asn	AGA Arg	ATC Ile	AAA Lys 100	TCA Ser	AAA Lys	CCA Pro	TTG Leu	GGT Gly 105	GAA Glu	AAG Lys	TTA Leu	TTA Leu	GAG Glu 110	ATG Met	ATT Ile	33	6
ATA Ile	AAT Asn	GGT Gly 115	ATA Ile	CCT Pro	TAT Tyr	CTT Leu	GGA Gly 120	GAT Asp	AGA Arg	CGT Arg	GTT Val	CCA Pro 125	CTC	GAA Glu	GAG Glu	38	4
TTT Phe	AAC Asn 130	ACA Thr	AAC Asn	ATT Ile	GCT Ala	AGT Ser 135	GTA Val	ACT Thr	GTT Val	AAT Asn	AAA Lys 140	TTA Leu	ATC Ile	AGT Ser	AAT Asn	43	2
Pro 145	Gly	GAA Glu	Val	Glu	Arg 150	Lys	Lys	Gly	Ile	Phe 155	Ala	Asn	Leu	Ile	Ile 160	480	0
TTT Phe	GGA Gly	Pro	GGG Gly	CCA Pro 165	GTT Val	TTA Leu	AAT Asn	GAA Glu	AAT Asn 170	GAG Glu	ACT Thr	ATA Ile	GAT Asp	ATA Ile 175	GGT Gly	528	3
ATA Ile	CAA Gln	AAT Asn	CAT ⁻ His 180	TTT Phe	GCA Ala	TCA Ser	AGG Arg	GAA Glu 185	GGC	TTC Phe	GGG Gly	GGT Gly	ATA Ile 190	ATG Met	CAA Gln	576	5
ATG Met	AAG Lys	TTT Phe 195	TGC Cys	CCA Pro	GAA Glu	TAT Tyr	GTA Val 200	AGC Ser	GTA Val	TTT Phe	AAT Asn	AAT Asn 205	GTT Val	CAA Gln	GAA Glu	624	,
AAC Asn	AAA Lys 210	GGC Gly	GCA Ala	AGT Ser	ATA Ile	TTT Phe 215	AAT Asn	AGA Arg	CGT Arg	Gly	TAT Tyr 220	TTT Phe	TCA Ser	GAT Asp	CCA Pro	672	?
GCC Ala 225	TTG Leu	ATA Ile-	TTA Leu -	ATG Met_	CAT His 230	GAA Glu	CTT Leu-	ATA Ile	CAT His-	GTT Va·l— 235	TTA Leu	CAT His	GGA Gly-	Leu-	TAT Tyr- 240	720	

- 96 -

	GGC Gly	ATT Ile	AAA Lys	GTA Val	GAT Asp 245	GAT Asp	TTA Leu	CCA Pro	ATT Ile	GTA Val 250	CCA Pro	AAT Asn	GAA Glu	AAA Lys	AAA Lys 255	TTT Phe	768
	TTT Phe	ATG Met	CAA Gln	TCT Ser 260	ACA Thr	GAT Asp	GCT Ala	ATA Ile	CAG Gln 265	GCA Ala	GAA Glu	GAA Glu	CTA Leu	TAT Tyr 270	ACA Thr	TTT Phe	816
	GGA Gly	GGA Gly	CAA Gln 275	GAT Asp	CCC Pro	AGC Ser	ATC Ile	ATA Ile 280	ACT Thr	CCT Pro	TCT Ser	ACG Thr	GAT Asp 285	AAA Lys	AGT Ser	ATC Ile	864
	TAT Tyr	GAT Asp 290	AAA Lys	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT Phe	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	GAT Asp	AGA Arg	CTT Leu	AAC Asn	912
	AAG Lys 305	GTT Val	TTA Leu	GTT Val	TGC Cys	ATA Ile 310	TCA Ser	GAT Asp	CCT Pro	AAC Asn	ATT Ile 315	AAT Asn	ATT Ile	AAT Asn	ATA Ile	TAT Tyr 320	960
	AAA Lys	AAT Asn	AAA Lys	TTT Phe	AAA Lys 325	GAT Asp	AAA Lys	TAT Tyr	AAA Lys	TTC Phe 330	GTT Val	GAA Glu	GAT Asp	TCT Ser	GAG Glu 335	GGA Gly	1008
	AAA Lys	TAT Tyr	AGT Ser	ATA Ile 340	GAT Asp	GTA Val	GAA Glu	AGT Ser	TTT Phe 345	GAT Asp	AAA Lys	TTA Leu	TAT Tyr	AAA Lys 350	AGC Ser	TTA Leu	1056
	ATG Met	TTT Phe	GGT Gly 355	TTT Phe	ACA Thr	GAA Glu	ACT Thr	AAT Asn 360	ATA Ile	GCA Ala	GAA Glu	AAT Asn	TAT Tyr 365	AAA Lys	ATA Ile	AAA Lys	1104
					TAT Tyr												1152
					AAT Asn												1200
					ATG Met 405												1248
					TAT Tyr												1296
					TGT Cys												1344
					GAT Asp												1392
					AAA Lys												1440
					GAC Asp 485												1488
-					ATA Ile												1536-

		AAT Asn 515													AAA Lys	•	1584
		TTT Phe															1632
		CCT Pro															1680
		TTA Leu															1728
		AAA Lys															1776
		AAA Lys 595															1824
		ATG Met															1872
		GCT Ala															1920
		TTT Phe															1968
		TTA Leu							Phe								2016
		AAA Lys 675															2064
		GAA Glu															2112
		ACA Thr															2160
		TTA Leu															2208
AGA Arg	TAT Tyr	TAA Asn	ATA Ile 740	TAT Tyr	TCT Ser	GAA Glu	AAA Lys	GAA Glu 745	AAG Lys	TCA Ser	AAT Asn	ATT Ile	AAC Asn 750	ATC Ile	GAT Asp		2256
		GAT Asp 755															2304
		ATA- Ile															2352

AAA Lys 785	AAA	ATG Met	ATT Ile	CCA Pro	TTA Leu 790	GCT Ala	GTA Val	GAA Glu	AAA Lys	TTA Leu 795	CTA Leu	GAC Asp	TTT Phe	GAT Asp	AAT Asn 800	2400
ACT Thr	CTC Leu	AAA Lys	AAA Lys	AAT Asn 805	TTG Leu	TTA Leu	AAT Asn	TAT Tyr	ATA Ile 810	GAT Asp	GAA Glu	AAT Asn	AAA Lys	TTA Leu 815	TAT Tyr	2448
TTG Leu	ATT Ile	GGA Gly	AGT Ser 820	GCA Ala	GAA Glu	TAT Tyr	GAA Glu	AAA Lys 825	TCA Ser	AAA Lys	GTA Val	AAT Asn	AAA Lys 830	TAC Tyr	TTG Leu	2496
AAA Lys	ACC Thr	ATT Ile 835	ATG Met	CCG Pro	TTT Phe	GAT Asp	CTT Leu 840	TCA Ser	ATA Ile	TAT Tyr	ACC Thr	AAT Asn 845	GAT Asp	ACA Thr	ATA Ile	2544
CTA Leu	ATA Ile 850	GAA Glu	ATG Met	TTT Phe	TAA neA	AAA Lys 855	TAT Tyr	AAT Asn	AGC Ser							2574

(2) INFORMATION FOR SEQ ID NO: 22:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 858 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22: Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Leu Glu Met Ile Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu Phe Asn Thr Asn Ile Ala Ser Val Thr Val Asn Lys Leu Ile Ser Asn Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile

Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln 180 185

Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly

Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Glu Glu 200 Asn Lys Gly Ala Ser Ile Phe Asn Arg Arg Gly Tyr Phe Ser Asp Pro Ala Leu Ile Leu Met His Glu Leu Ile His Val Leu His Gly Leu Tyr 230 Gly Ile Lys Val Asp Asp Leu Pro Ile Val Pro Asn Glu Lys Lys Phe Phe Met Gln Ser Thr Asp Ala Ile Gln Ala Glu Glu Leu Tyr Thr Phe Gly Gln Asp Pro Ser Ile Ile Thr Pro Ser Thr Asp Lys Ser Ile 280 Tyr Asp Lys Val Leu Gln Asn Phe Arg Gly Ile Val Asp Arg Leu Asn 295 Lys Val Leu Val Cys Ile Ser Asp Pro Asn Ile Asn Ile Asn Ile Tyr Lys Asn Lys Phe Lys Asp Lys Tyr Lys Phe Val Glu Asp Ser Glu Gly 330 Lys Tyr Ser Ile Asp Val Glu Ser Phe Asp Lys Leu Tyr Lys Ser Leu 345 Met Phe Gly Phe Thr Glu Thr Asn Ile Ala Glu Asn Tyr Lys Ile Lys Thr Arg Ala Ser Tyr Phe Ser Asp Ser Leu Pro Pro Val Lys Ile Lys Asn Leu Leu Asp Asn Glu Ile Tyr Thr Ile Glu Glu Gly Phe Asn Ile Ser Asp Lys Asp Met Glu Lys Glu Tyr Arg Gly Gln Asn Lys Ala Ile 410 Asn Lys Gln Ala Tyr Glu Glu Ile Ser Lys Glu His Leu Ala Val Tyr Lys Ile Gln Met Cys Lys Ser Val Lys Ala Pro Gly Ile Cys Ile Asp Val Asp Asn Glu Asp Leu Phe Phe Ile Ala Asp Lys Asn Ser Phe Ser 455 Asp Asp Leu Ser Lys Asn Glu Arg Ile Glu Tyr Asn Thr Gln Ser Asn 470 Tyr Ile Glu Asn Asp Phe Pro Ile Asn Glu Leu Ile Leu Asp Thr Asp 485 490 Leu Ile Ser Lys Ile Glu Leu Pro Ser Glu Asn Thr Glu Ser Leu Thr Asp Phe Asn Val Asp Val Pro Val Tyr Glu Lys Gln Pro Ala Ile Lys Lys Ile Phe Thr Asp Glu Asn Thr Ile Phe Gln Tyr Leu Tyr Ser Gln ___ 535-- ---- -- -540----

Thr Phe Pro Leu Asp Ile Arg Asp Ile Ser Leu Thr Ser Ser Phe Asp Asp Ala Leu Leu Phe Ser Asn Lys Val Tyr Ser Phe Phe Ser Met Asp Tyr Ile Lys Thr Ala Asn Lys Val Val Glu Ala Gly Leu Phe Ala Gly Trp Val Lys Gln Ile Val Asn Asp Phe Val Ile Glu Ala Asn Lys Ser Asn Thr Met Asp Lys Ile Ala Asp Ile Ser Leu Ile Val Pro Tyr Ile Gly Leu Ala Leu Asn Val Gly Asn Glu Thr Ala Lys Gly Asn Phe Glu 630 Asn Ala Phe Glu Ile Ala Gly Ala Ser Ile Leu Leu Glu Phe Ile Pro Glu Leu Leu Ile Pro Val Val Gly Ala Phe Leu Leu Glu Ser Tyr Ile Asp Asn Lys Asn Lys Ile Ile Lys Thr Ile Asp Asn Ala Leu Thr Lys 680 Arg Asn Glu Lys Trp Ser Asp Met Tyr Gly Leu Ile Val Ala Gln Trp Leu Ser Thr Val Asn Thr Gln Phe Tyr Thr Ile Lys Glu Gly Met Tyr Lys Ala Leu Asn Tyr Gln Ala Gln Ala Leu Glu Glu Ile Ile Lys Tyr 730 Arg Tyr Asn Ile Tyr Ser Glu Lys Glu Lys Ser Asn Ile Asn Ile Asp Phe Asn Asp Ile Asn Ser Lys Leu Asn Glu Gly Ile Asn Gln Ala Ile Asp Asn Ile Asn Asn Phe Ile Asn Gly Cys Ser Val Ser Tyr Leu Met Lys Lys Met Ile Pro Leu Ala Val Glu Lys Leu Leu Asp Phe Asp Asn Thr Leu Lys Lys Asn Leu Leu Asn Tyr Ile Asp Glu Asn Lys Leu Tyr Leu Ile Gly Ser Ala Glu Tyr Glu Lys Ser Lys Val Asn Lys Tyr Leu 825 Lys Thr Ile Met Pro Phe Asp Leu Ser Ile Tyr Thr Asn Asp Thr Ile Leu Ile Glu Met Phe Asn Lys Tyr Asn Ser

(2) INFORMATION FOR SEQ ID NO: 23:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1644 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

PCT/GB97/02273

- 101 -

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION:1..1644

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

ATG Met 1	CCA Pro	GTT Val	ACA Thr	ATA Ile 5	AAT Asn	AAT Asn	TTT Phe	AAT Asn	TAT Tyr 10	AAT Asn	GAT Asp	CCT Pro	ATT	GAT Asp 15	AAT Asn	••	48
AAT Asn	AAT Asn	ATT Ile	ATT Ile 20	ATG Met	ATG Met	GAG Glu	CCT Pro	CCA Pro 25	TTT Phe	GCG Ala	AGA Arg	GGT Gly	ACG Thr 30	GGG Gly	AGA Arg		96
											TGG Trp						144
											AAT Asn 60						192
											CCA Pro						240
											ATG Met						288
											TTA Leu						336
											GTT Val						384
											AAA Lys 140						432
											GCA Ala						480
											ACT Thr						528
											GGG Gly						576
											AAT Asn						624
											TAT Tyr 220						672

- 102 -

GCC Ala 225	TTG Leu	ATA Ile	TTA Leu	ATG Met	CAT His 230	GAA Glu	CTT Leu	ATA Ile	CAT His	GTT Val 235	TTA Leu	CAT His	GGA Gly	TTA Leu	TAT Tyr 240		720
GGC Gly	ATT Ile	AAA Lys	GTA Val	GAT Asp 245	GAT Asp	TTA Leu	CCA Pro	ATT Ile	GTA Val 250	CCA Pro	AAT Asn	GAA Glu	AAA Lys	AAA Lys 255	TTT Phe		768
TTT Phe	ATG Met	CAA Gln	TCT Ser 260	ACA Thr	GAT Asp	GCT Ala	ATA Ile	CAG Gln 265	GCA Ala	GAA Glu	GAA Glu	CTA Leu	TAT Tyr 270	ACA Thr	TTT Phe		816
				CCC Pro													864
TAT Tyr	GAT Asp 290	AAA Lys	GTT Val	TTG Leu	CAA Gln	AAT Asn 295	TTT Phe	AGA Arg	GGG Gly	ATA Ile	GTT Val 300	GAT Asp	AGA Arg	CTT Leu	AAC Asn		912
				TGC Cys													960
				AAA Lys 325													1008
				GAT Asp													1056
				ACA Thr													1104
				TAT Tyr													1152
AAT Asn 385	TTA Leu	TTA Leu	GAT Asp	AAT Asn	GAA Glu 390	ATC Ile	TAT Tyr	ACT Thr	ATA Ile	GAG Glu 395	GAA Glu	GGG Gly	TTT Phe	AAT Asn	ATA Ile 400		1200
				ATG Met 405	Glu	Lys	Glu	Tyr	Arg	Gly	Gln	Asn		Ala	Ile		1248
				TAT Tyr													1296
				TGT Cys													1344
				GAT Asp												-	1392
				AAA Lys													1440
 TAT	ATA Ile	GAA Glu	AAT Asn	GAC Asp 485	TTC Phe	CCT Pro	ATA Ile	AAT Asn	GAA Glu 490	TTA Leu	ATT Ile	TTA Leu	GAT Asp	ACT Thr 495	GAT Asp		1488

PCT/GB97/02273

- 103 -

TTA Leu	ATA Ile	AGT Ser	AAA Lys 500	ATA Ile	GAA Glu	TTA Leu	CCA Pro	AGT Ser 505	GAA Glu	AAT Asn	ACA Thr	GAA Glu	TCA Ser 510	CTT Leu	ACT Thr	٠	1536
GAT Asp	TTT Phe	AAT Asn 515	GTA Val	GAT Asp	GTT Val	CCA Pro	GTA Val 520	TAT Tyr	GAA Glu	AAA Lys	CAA Gln	CCC Pro 525	GCT Ala	ATA Ile	AAA Lys		1584
						AAT Asn 535											1632
	TTT Phe																1644

(2) INFORMATION FOR SEQ ID NO: 24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 548 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

Met Pro Val Thr Ile Asn Asn Phe Asn Tyr Asn Asp Pro Ile Asp Asn 1 5 10 15 Asn Asn Ile Ile Met Met Glu Pro Pro Phe Ala Arg Gly Thr Gly Arg Tyr Tyr Lys Ala Phe Lys Ile Thr Asp Arg Ile Trp Ile Ile Pro Glu 35 40 45 Arg Tyr Thr Phe Gly Tyr Lys Pro Glu Asp Phe Asn Lys Ser Ser Gly 50 55 Ile Phe Asn Arg Asp Val Cys Glu Tyr Tyr Asp Pro Asp Tyr Leu Asn 65 70 75 80 Thr Asn Asp Lys Lys Asn Ile Phe Leu Gln Thr Met Ile Lys Leu Phe Asn Arg Ile Lys Ser Lys Pro Leu Gly Glu Lys Leu Glu Met Ile 100 105 110 Ile Asn Gly Ile Pro Tyr Leu Gly Asp Arg Arg Val Pro Leu Glu Glu Phe Asn Thr Asn Ile Ala Ser Val Thr Val As $f \pi^-$ Lys Leu Ile Ser Asn Pro Gly Glu Val Glu Arg Lys Lys Gly Ile Phe Ala Asn Leu Ile Ile 155 Phe Gly Pro Gly Pro Val Leu Asn Glu Asn Glu Thr Ile Asp Ile Gly 165 Ile Gln Asn His Phe Ala Ser Arg Glu Gly Phe Gly Gly Ile Met Gln 185

Met Lys Phe Cys Pro Glu Tyr Val Ser Val Phe Asn Asn Val Gln Glu

										10	Τ -				
Asn	Lys 210	Gly	/ Ala	Ser	·Ile	Phe 215	Asn	Arg	Arg	Gly	Tyr 220		Ser	Asp	Pro
Ala 225	Leu	Ile	Leu	Met	His 230	Glu	Leu	Ile	His	Val 235	Leu	His	Gly	Leu	Tyr 240
Gly	Ile	Lys	Val	Asp 245	Asp	Leu	Pro	Ile	Val 250	Pro	Asn	Glu	Lys	Lys 255	Phe
Phe	Met	Gln	Ser 260	Thr	Asp	Ala	Ile	Gln 265	Ala	Glu	Glu	Leu	Tyr 270	Thr	Phe
		2/5					280					285		Ser	
	290					295					300			Leu	
305					310					315				Ile	320
				325					330					Glu 335	-
			340					345					350	Ser	
		355					360					365		Ile	
	370					375					380			Ile	_
385					390					395				Asn	400
				405					410					Ala 415	
			420					425					430	Val	•
		435	•				440					445		Ile	_
	450					455					460			Phe	
465					470					475				Ser	480
				485					490					Thr 495	-
			500					505					510	Leu	
		515					520					525		Ile	
Lys	Ile 530	Phe	Thr	Asp	Glu	Asn 535	Thr	Ile	Phe	Gln	Tyr 540	Leu	Tyr	Ser	Gln
Thr	Phe	Pro	Leu												

WO 98/07864 PCT/GB97/02273

- 105 -

(2) INFORMATION FOR SEQ ID NO: 25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2616 base pairs

 - (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (ix) FEATURE:

 - (A) NAME/KEY: CDS
 (B) LOCATION:1..2616

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

ATG CAG TTC GTG AAC AAG CAG TTC AAC TAT AAG GAC CCT GTA AAC GGT Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly 1 S S S Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly 1 S S S Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly GTG GAC ATT GCC TAC ATC AAA ATT CCA AAC GCC GGC CAG ATG CAG CCG Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 S 10 GTG AAG GCT TTC AAG ATT CAT AAC AAA ATC TGG GTT ATT CCG GAA CGC Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg 35 40 GAT ACA TTT ACG AAC CCG GAA GAA GGA GAC TTG AAC CCG CCG CGG GAA Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 S S S S S S S S S S S S S S S S S S S																	
Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro 20 25 30 30 30 30 30 30 30 30 30 30 30 30 30	ATG Met 1	CAG Gln	TTC Phe	GTG Val	AAC Asn 5	AAG Lys	CAG Gln	TTC Phe	AAC Asn	Tyr	AAG Lys	GAC Asp	CCT Pro	GTA Val	Asn	GGT Gly	48
Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg GAT ACA TTT ACG AAC CCG GAA GAA GGA GGA GAC TTG AAC CCG CCG CCG GAA Asp Thr Phe Thr Asn Pro Glu Glu Glu Asp Leu Asn Pro Pro Pro Glu 50 GCA AAG CAG GTG CCA GTT TCA TAC TAC TAC GAT TCA ACC TAT CTG AGC ACA Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 GAC AAC GAG AAG GAT AAC TC CTG AAG GGA GTG ACC AAA TTA TTC GAG Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 95 CGT ATT TAT TCC ACT GAC GGC GGC CGT ATG CTG CTG ACC TCA ATC GTC Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 CGT ACT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 CGT ACG TTT GAG TCC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 175 CCGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CTT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe	GTT Val	GAC Asp	ATT Ile	Ala	TAC Tyr	ATC Ile	AAA Lys	ATT Ile	Pro	AAC Asn	GCC Ala	GGC Gly	CAG Gln	Met	CAG Gln	CCG Pro	96
Asp Thr Phe Thr Asn Pro Glu Glu Glu Asp Leu Asn Pro Pro Pro Glu GCA AAG CAG GTG CCA GTT TCA TAC TAC TAC ACC TAT CTG AGC ACA Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65	GTG Val	AAG Lys	Ala	TTC Phe	AAG Lys	ATT Ile	CAT His	Asn	AAA Lys	ATC Ile	TGG Trp	GTT Val	Ile	CCG Pro	GAA Glu	CGC Arg	144
Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 80 GAC AAC GAG AAG GAT AAC TAC CTG AAG GGA GTG ACC AAA TTA TTC GAG ASp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 95 CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA ATC GTC Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 110 CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 125 GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Val Ile Ile Glp Pro Ser Ala Asp Ile 150 ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTA ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 175 CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe	GAT Asp	Thr	TTT Phe	ACG Thr	AAC Asn	CCG Pro	Glu	GAA Glu	GGA Gly	GAC Asp	TTG Leu	Asn	CCG Pro	CCG Pro	CCG Pro	GAA Glu	192
Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 95 CGT ATT TAT TCC ACT GAC CTG GGC CGT ATG CTG CTG ACC TCA ATC GTC Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 110 CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 125 GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 175 CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe	Ala	AAG Lys	CAG Gln	GTG Val	CCA Pro	Val	TCA Ser	TAC Tyr	TAC Tyr	GAT Asp	Ser	ACC Thr	TAT Tyr	CTG Leu	AGC Ser	Thr	240
Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val 100 CGC GGA ATC CCA TTT TGG GGT GGC AGT ACC ATT GAC ACG GAG TTG AAG Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe					Asp					Gly					Phe		288
Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys 115 GTT ATT GAC ACT AAC TGC ATT AAC GTG ATC CAA CCA GAC GGT AGC TAC Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe				Ser					Arg					Ser			336
Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 130 AGA TCT GAA GAA CTT AAC CTC GTA ATC ATC GGG CCC TCC GCG GAC ATT Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe	CGC Arg	GGA Gly	Ile	CCA Pro	TTT Phe	TGG Trp	GGT Gly	Gly	AGT Ser	ACC Thr	ATT Ile	GAC Asp	Thr	GAG Glu	TTG Leu	AAG Lys	384
Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile 145 ATC CAG TTT GAG TGC AAG AGC TTT GGC CAC GAA GTG TTG AAC CTG ACG Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe	GTT Val	Ile	GAC Asp	ACT Thr	AAC Asn	TGC Cys	Ile	AAC Asn	GTG Val	ATC Ile	CAA Gln	Pro	GAC Asp	GGT Gly	AGC Ser	TAC Tyr	432
Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr 165 170 175 CGT AAC GGT TAC GGC TCT ACT CAG TAC ATT CGT TTC AGC CCA GAC TTC Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe	Arg					Asn					Gly					Ile	480
Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe				Glu	Суѕ					His					Leu		528
	CGT Arg	AAC Asn	GGT Gly	Tyr	GGC Gly	TCT Ser	ACT Thr	CAG Gln	Tyr	ATT Ile	CGT Arg	TTC Phe	AGC Ser	Pro	GAC Asp	TTC Phe	 576

- 106 -

				GAG Glu												624
GGT Gly	GCA Ala 210	GGC Gly	AAG Lys	TTC Phe	GCA Ala	ACT Thr 215	GAT Asp	CCA Pro	GCG Ala	GTG Val	ACC Thr 220	CTG Leu	GCA Ala	CAC His	GAG Glu	672
CTG Leu 225	ATC Ile	CAC His	GCC Ala	GGT Gly	CAT His 230	CGT Arg	CTG Leu	TAT	GGC Gly	ATT Ile 235	GCG Ala	ATT Ile	AAC Asn	CCG Pro	AAC Asn 240	720
CGC Arg	GTG Val	TTC Phe	AAG Lys	GTT Val 245	AAC Asn	ACC Thr	AAC Asn	GCC Ala	TAC Tyr 250	TAC Tyr	GAG Glu	ATG Met	AGT Ser	GGT Gly 255	TTA Leu	768
				GAG Glu												816
				TTG Leu												864
				ATT Ile												912
				TCA Ser												960
				GAA Glu 325												1008
				TTA Leu												1056
				TTT Phe												1104
Phe	GAT Asp 370	Lys	GCC Ala	GTA Val	TTT Phe	AAG Lys 375	ATA Ile	AAT Asn	ATA Ile	GTA Val	CCT Pro 380	AAG Lys	GTA Val	AAT Asn	TAC Tyr	1152
ACA Thr 385	ATA Ile	TAT Tyr	GAT Asp	GGA Gly	TTT Phe 390	ÁAT Asn	TTA Leu	AGA Arg	AAT Asn	ACA Thr 395	AAT Asn	TTA Leu	GCA Ala	GCA Ala	AAC Asn 400	1200
TTT Phe	AAT Asn	GGT Gly	CAA Gln	AAT Asn 405	ACA Thr	GAA Glu	ATT Ile	AAT Asn	AAT Asn 410	ATG Met	TAA Ran	TTT Phe	ACT Thr	AAA Lys 415	CTA Leu	1248
AAA Lys	AAT Asn	TTT Phe	ACT Thr 420	GGA Gly	TTG Leu	TTT Phe	GAA Glu	TTT Phe 425	TAT Tyr	AAG Lys	TTG Leu	CTA Leu	TGT Cys 430	GTA Val	AGA Arg	1296
GGG Gly	ATA Ile	ATA Ile 435	ACT Thr	TCT Ser	AAA Lys	ACT Thr	AAA Lys 440	TCA Ser	TTA Leu	GAT Asp	AAA Lys	GGA Gly 445	TAC Tyr	AAT Asn	AAG Lys	1344
GCA Ala	TTA Leu 450	AAT Asn	GAT Asp	TTA Leu	TGT Cys	ATC Ile 455	AAA Lys	GTT Val	AAT Asn	`AAT Asn	TGG Trp 460	GAC Asp	TTG Leu	TTT Phe	TTT Phe	1392

AGT Ser 465	Pro	TCA Ser	GAA Glu	GAT Asp	AAT Asn 470	Phe	ACT Thr	AAT Asn	GAT Asp	CTA Leu 475	Asn	AAA Lys	GGA Gly	GAA Glu	GAA Glu 480	1440
ATT Ile	ACA Thr	TCT Ser	GAT Asp	ACT Thr 485	Asn	ATA Ile	GAA Glu	GCA Ala	GCA Ala 490	GAA Glu	GAA Glu	AAT Asn	ATT Ile	AGT Ser 495	TTA Leu	1488
GAT Asp	TTA Leu	ATA Ile	CAA Gln 500	CAA Gln	TAT	TAT	TTA Leu	ACC Thr 505	Phe	AAT Asn	TTT Phe	GAT Asp	AAT Asn 510	Glu	CCT	1536
Glu	Asn	Ile 515	Ser	Ile	Glu	Asn	Leu 520	Ser	Ser	Asp	Ile	Ile 525	Gly	Gln	TTA Leu	1584
GAA Glu	CTT Leu 530	ATG Met	CCT Pro	AAT Asn	ATA Ile	GAA Glu 535	AGA Arg	TTT	CCT Pro	AAT Asn	GGA Gly 540	AAA Lys	AAG Lys	TAT	GAG Glu	1632
TTA Leu 545	GAT Asp	AAA Lys	TAT Tyr	ACT Thr	ATG Met 550	TTC Phe	CAT His	TAT	CTT	CGT Arg 555	GCT Ala	CAA Gln	GAA Glu	TTT Phe	GAA Glu 560	1680
CAT His	GGT Gly	AAA Lys	TCT Ser	AGG Arg 565	ATT Ile	GCT Ala	TTA Leu	ACA Thr	AAT Asn 570	TCT Ser	GTT Val	AAC Asn	GAA Glu	GCA Ala 575	TTA Leu	1728
TTA Leu	AAT Asn	CCT Pro	AGT Ser 580	CGT Arg	GTT Val	TAT Tyr	ACA Thr	TTT Phe 585	TTT Phe	TCT Ser	TCA Ser	GAC Asp	TAT Tyr 590	GTA Val	AAG Lys	1776
AAA Lys	GTT Val	AAT Asn 595	AAA Lys	GCT Ala	ACG Thr	GAG Glu	GCA Ala 600	GCT Ala	ATG Met	TTT Phe	TTA Leu	GGC Gly 605	TGG Trp	GTA Val	GAA Glu	1824
CAA Gln	TTA Leu 610	GTA Val	TAT Tyr	GAT Asp	TTT Phe	ACC Thr 615	GAT Asp	GAA Glu	ACT Thr	AGC Ser	GAA Glu 620	GTA Val	AGT Ser	ACT Thr	ACG Thr	1872
GAT Asp 625	AAA Lys	ATT Ile	GCG Ala	GAT Asp	ATA Ile 630	ACT Thr	ATA Ile	ATT Ile	ATT Ile	CCA Pro 635	TAT Tyr	ATA Ile	GGA Gly	CCT Pro	GCT Ala 640	1920
TTA Leu	TAA Asn	ATA Ile	GGT Gly	AAT Asn 645	ATG Met	TTA Leu	TAT Tyr	AAA Lys	GAT Asp 650	GAT Asp	TTT Phe	GTA Val	GGT Gly	GCT Ala 655	TTA Leu	1968
ATA Ile	TTT Phe	TCA Ser	GGA Gly 660	GCT Ala	GTT Val	ATT Ile	CTG Leu	TTA Leu 665	GAA Glu	TTT Phe	ATA Ile	CCA Pro	GAG Glu 670	ATT Ile	GCA Ala	2016
ATA Ile	CCT Pro	GTA Val 675	TTA Leu	GGT Gly	ACT Thr	TTT Phe	GCA Ala 680	CTT Leu	GTA Val	TCA Ser	TAT Tyr	ATT Ile 685	GCG Ala	AAT Asn	AAG Lys	2064
GTT Val	CTA Leu 690	ACC Thr	GTT Val	CAA Gln	ACA Thr	ATA Ile 695	GAT Asp	AAT Asn	GCT Ala	TTA Leu	AGT Ser 700	AAA Lys	AGA Arg	AAT Asn	GAA Glu	2112
AAA Lys 705	TGG Trp	GAT Asp	GAG Glu	GTC Val	TAT Tyr 710	AAA Lys	TAT Tyr	ATA Ile	GTA Val	ACA Thr 715	AAT Asn	TGG Trp	TTA Leu	GCA Ala	AAG Lys 720	2160
GTT Val	AAT Asn	ACA_ Thr	CAG Gln	ATT Ile 725	GAT. Asp	.CTA. Leu	ATA Ile	AGA- Arg	AAA Lys 730	AAA Lys	ATG - Met	AAA Lys	GAA- Glu	GCT Ala 735	TTA- Leu	2208

- 108 -

GAA Glu	AAT Asn	CAA Gln	GCA Ala 740	GAA Glu	GCA Ala	ACA Thr	AAG Lys	GCT Ala 745	ATA Ile	ATA Ile	AAC Asn	TAT Tyr	CAG Gln 750	TAT Tyr	AAT Asn		2256
					GAG Glu												2304
					AAT Asn												2352
					CAA Gln 790												2400
					AAA Lys												2448
					TAT Tyr						-					•	2496
					AAA Lys												2544
					TCC Ser												2592
					ATT Ile 870		TAA •										2616

(2) INFORMATION FOR SEQ ID NO: 26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 872 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Met Gln Phe Val Asn Lys Gln Phe Asn Tyr Lys Asp Pro Val Asn Gly

Val Asp Ile Ala Tyr Ile Lys Ile Pro Asn Ala Gly Gln Met Gln Pro

Val Lys Ala Phe Lys Ile His Asn Lys Ile Trp Val Ile Pro Glu Arg
35 40 45

Asp Thr Phe Thr Asn Pro Glu Glu Gly Asp Leu Asn Pro Pro Pro Glu 50 55 60

Ala Lys Gln Val Pro Val Ser Tyr Tyr Asp Ser Thr Tyr Leu Ser Thr 65 70 75 80

Asp Asn Glu Lys Asp Asn Tyr Leu Lys Gly Val Thr Lys Leu Phe Glu 85 90 95

Arg Ile Tyr Ser Thr Asp Leu Gly Arg Met Leu Leu Thr Ser Ile Val

WO 98/07864 PCT/GB97/02273

- 109 -

Arg Gly Ile Pro Phe Trp Gly Gly Ser Thr Ile Asp Thr Glu Leu Lys Val Ile Asp Thr Asn Cys Ile Asn Val Ile Gln Pro Asp Gly Ser Tyr 135 Arg Ser Glu Glu Leu Asn Leu Val Ile Ile Gly Pro Ser Ala Asp Ile Ile Gln Phe Glu Cys Lys Ser Phe Gly His Glu Val Leu Asn Leu Thr Arg Asn Gly Tyr Gly Ser Thr Gln Tyr Ile Arg Phe Ser Pro Asp Phe Thr Phe Gly Phe Glu Glu Ser Leu Glu Val Asp Thr Asn Pro Leu Leu 200 Gly Ala Gly Lys Phe Ala Thr Asp Pro Ala Val Thr Leu Ala His Glu 215 Leu Ile His Ala Gly His Arg Leu Tyr Gly Ile Ala Ile Asn Pro Asn 230 Arg Val Phe Lys Val Asn Thr Asn Ala Tyr Tyr Glu Met Ser Gly Leu Glu Val Ser Phe Glu Glu Leu Arg Thr Phe Gly Gly His Asp Ala Lys 265 Phe Ile Asp Ser Leu Gln Glu Asn Glu Phe Arg Leu Tyr Tyr Asn 280 Lys Phe Lys Asp Ile Ala Ser Thr Leu Asn Lys Ala Lys Ser Ile Val Gly Thr Thr Ala Ser Leu Gln Tyr Met Lys Asn Val Phe Lys Glu Lys Tyr Leu Leu Ser Glu Asp Thr Ser Gly Lys Phe Ser Val Asp Lys Leu Lys Phe Asp Lys Leu Tyr Lys Met Leu Thr Glu Ile Tyr Thr Glu Asp 345 Asn Phe Val Lys Phe Phe Lys Val Leu Asn Arg Lys Thr Tyr Leu Asn 360 Phe Asp Lys Ala Val Phe Lys Ile Asn Ile Val Pro Lys Val Asn Tyr Thr Ile Tyr Asp Gly Phe Asn Leu Arg Asn Thr Asn Leu Ala Ala Asn 395 Phe Asn Gly Gln Asn Thr Glu Ile Asn Asn Met Asn Phe Thr Lys Leu 410 Lys Asn Phe Thr Gly Leu Phe Glu Phe Tyr Lys Leu Leu Cys Val Arg Gly Ile Ile Thr Ser Lys Thr Lys Ser Leu Asp Lys Gly Tyr Asn Lys Ala Leu Asn Asp Leu Cys Ile Lys Val Asn Asn Trp Asp Leu Phe Phe _____455

- 110 -

Ser 465	Pro	Ser	Glu	Asp	Asn 470	Phe	Thr	Asn	Asp	Leu 475	Asn	Lys	Gly	Glu	Glu 480
Ile	Thr	Ser	Asp	Thr 485	Asn	Ile	Glu	Ala	Ala 490	Glu	Glu	Asn	Ile	Ser 495	Leu
Asp	Leu	Ile	Gln 500	Gln	Tyr	Tyr	Leu	Thr 505	Phe	Asn	Phe	Asp	Asn 510	Glu	Pro
Glu	Asn	Ile 515	Ser	Ile	Glu	Asn	Leu 520	Ser	Ser	Asp	Ile	Ile 525	Gly	Gln	Leu
Glu	Leu 530	Met	Pro	Asn	Ile	Glu 535	Arg	Phe	Pro	Asn	Gly 540	Lys	Lys	Tyr	Glu
Leu 545	Asp	Lys	Tyr	Thr	Met 550	Phe	His	Tyr	Leu	Arg 555	Ala	Gln	Glu	Phe	Glu 560
His	Gly	Lys	Ser	Arg 565	Ile	Ala	Leu	Thr	Asn 570	Ser	Val	Asn	Glu	Ala 575	Leu
Leu	Asn	Pro	Ser 580	Arg	Val	Tyr	Thr	Phe 585	Phe	Ser	Ser	Asp	Tyr 590	Val	Lys
Lys	Val	Asn 595	Lys	Ala	Thr	Glu	Ala 600	Ala	Met	Phe	Leu	Gly 605	Trp	Val	Glu
Gln	Leu 610	Val	Tyr	Asp	Phe	Thr 615	Asp	Glu	Thr	Ser	Glu 620	Val	Ser	Thr	Thr
Asp 625	ГÀг	Ile	Ala	Asp	Ile 630	Thr	Ile	Ile	Ile	Pro 635	Туг	Ile	Gly	Pro	Ala 640
Leu	Asn	Ile	Gly	Asn 645	Met	Leu	Tyr	Lys	Asp 650	Asp	Phe	Val	Gly	Ala 655	Leu
Ile	Phe	Ser	Gly 660	Ala	Val	Ile	Leu	Leu 665	Glu	Phe	Ile	Pro	Glu 670	Ile	Ala
Ile	Pro	Val 675	Leu	Gly	Thr	Phe	Ala 680	Leu	Val	Ser	Tyr	Ile 685	Ala	Asn	Lys
Val	Leu 690	Thr	Val	Gln	Thr	Ile 695	Asp	Asn	Ala	Leu	Ser 700	Lys	Arg	Asn	Glu
Lys 705	Trp	Asp	Glu	Val	Tyr 710	Lys	Tyr	Ile	Val	Thr 715	Asn	Trp	Leu	Ala	Lys 720
Val	Asn	Thr	Gln	11e 725	Asp	Leu	Ile	Arg	Lys 730	Lys	Met	Lys	Glu	Ala 735	Leu
Glu	Asn	Gln	Ala 740	Glu	Ala	Thr	Lys	Ala 745	Ile	Ile	Asn	Tyr	Gln 750	Tyr	Asn
Gln	Tyr	Thr 755	Glu	Glu	Glu	Lys	Asn 760	Asn	Ile	Asn	Phe	Asn 765	Ile	Asp	Ąsp
Leu	Ser 770	Ser	Lys	Leu	Asn	Glu 775	Ser	Ile	Asn	Lys	Ala 780	Met	Ile	Asn	Ile
Asn 785	Lys	Phe	Leu	Asn	Gln 790	Cys	Ser	Val	Ser	Tyr 795	Leu	Met	Asn	Ser	Met 800
Ile	Pro	Tyr	Gly	Val -805		Arg	Leu		Asp -810					Leu -815	

WO 98/07864 PCT/GB97/02273

- 111 -

Asp Ala Leu Leu Lys Tyr Ile Tyr Asp Asn Arg Gly Thr Leu Ile Gly 825

Gln Val Asp Arg Leu Lys Asp Lys Val Asn Asn Thr Leu Ser Thr Asp 840 845

Ile Pro Phe Gln Leu Ser Lys Tyr Val Asp Asn Gln Arg Leu Leu Ser 860

Thr Phe Thr Glu Tyr Ile Lys *

- (2) INFORMATION FOR SEQ ID NO: 27:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2574 base pairs

 - (B) TYPE: nucleic acid (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (ii) MOLECULE TYPE: DNA (genomic)
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

ATGCCGGTTA	CCATCAACAA	CTTCAACTAC	AACGACCCGA	TCGACAACAA	CAACATCATC	60
ATGATGGAAC	CGCCGTTCGC	ACGTGGTACC	GGTCGTTACT	ACAAGGCTTT	CAAGATCACC	120
GACCGTATCT	GGATCATCCC	GGAACGTTAC	ACCTTCGGTT	ACAAACCTGA	GGACTTCAAC	180
AAGAGTAGCG	GGATTTTCAA	TCGTGACGTC	TGCGAGTACT	ATGATCCAGA	TTATCTGAAT	240
ACCAACGATA	AGAAGAACAT	ATTCCTTCAG	ACTATGATCA	AGTTATTTAA	TAGAATCAAA	300
TCAAAACCAT	TGGGTGAAAA	GTTATTAGAG	ATGATTATAA	ATGGTATACC	TTATCTTGGA	360
GATAGACGTG	TTCCACTCGA	AGAGTTTAAC	ACAAACATTG	CTAGTGTAAC	TGTTAATAAA	420
TTAATCAGTA	ATCCAGGAGA	AGTGGAGCGA	AAAAAAGGTA	TTTTCGCAAA	TTTAATAATA	480
TTTGGACCTG	GGCCAGTTTT	AAATGAAAAT	GAGACTATAG	ATATAGGTAT	ACAAAATCAT	540
TTTGCATCAA	GGGAAGGCTT	CGGGGGTATA	ATGCAAATGA	AGTTTTGCCC	AGAATATGTA	600
AGCGTATTTA	ATAATGTTCA	AGAAAACAAA	GGCGCAAGTA	TATTTAATAG	ACGTGGATAT	660
TTTTCAGATC	CAGCCTTGAT	ATTAATGCAT	GAACTTATAC	ATGTTTTACA	TGGATTATAT	720
GGCATTAAAG	TAGATGATTT	ACCAATTGTA	CCAAATGAAA	AAAAATTTTT	TATGCAATCT	780
ACAGATGCTA	TACAGGCAGA	AGAACTATAT	ACATTTGGAG	GACAAGATCC	CAGCATCATA	840
ACTCCTTCTA	CGGATAAAAG	TATCTATGAT	AAAGTTTTGC	AAAATTTTAG	AGGGATAGTT	900
GATAGACTTA	ACAAGGTTTT	AGTTTGCATA	TCAGATCCTA	ACATTAATAT	TAATATATAT	960
AAAAATAAAT	TTAAAGATAA	ATATAAATTC	GTTGAAGATT	CTGAGGGAAA	ATATAGTATA	1020
GATGTAGAAA	GTTTTGATAA	ATTATATAAA	AGCTTAATGT	TTGGTTTTAC	AGAAACTAAT	1080
ATAGCAGAAA	ATTATAAAAT	AAAAACTAGA	GCTTCTTATT	TTAGTGATTC	CTTACCACCA	1140
GTAAAAATAA	TTATTTAAAA	AGATAATGAA	ATCTATACTA	TAGAGGAAGG	GTTTAATATA	1200

- 112 -

TCTGAT	ĀAAG	ATATGGAAAA	AGAATATAGA	GGTCAGAATA	AAGCTATAAA	TAAACAAGCT	1260
TATGAA	GAAA	TTAGCAAGGA	GCATTTGGCT	GTATATAAGA	TACAAATGTG	TAAAAGTGTT	1320
AAAGCT	CCAG	GAATATGTAT	TGATGTTGAT	AATGAAGATT	TGTTCTTTAT	AGCTGATAAA	1380
AATAGT	TTTT	CAGATGATTT	ATCTAAAAAC	GAAAGAATAG	AATATAATAC	ACAGAGTAAT	1440
TATATAC	GAAA	ATGACTTCCC	TATAAATGAA	TTAATTTTAG	ATACTGATTT	AATAAGTAAA	1500
ATAGAA:	TAC	CAAGTGAAAA	TACAGAATCA	CTTACTGATT	TTAATGTAGA	TGTTCCAGTA	1560
TATGAA	AAAC	AACCCGCTAT	AAAAAAATT	TTTACAGATG	AAAATACCAT	CTTTCAATAT	1620
TTATACT	rctc	AGACATTTCC	TCTAGATATA	AGAGATATAA	GTTTAACATC	TTCATTTGAT	1680
GATGCAT	TAT	TATTTTCTAA	CAAAGTTTAT	TCATTTTTT	CTATGGATTA	TATTAAAACT	1740
GCTAAT	AAAG	TGGTAGAAGC	AGGATTATTT	GCAGGTTGGG	TGAAACAGAT	AGTAAATGAT	1800
TTTGTAA	ATCG	AAGCTAATAA	AAGCAATACT	ATGGATAAAA	TTGCAGATAT	ATCTCTAATT	1860
GTTCCTT	ATA	TAGGATTAGC	TTTAAATGTA	GGAAATGAAA	CAGCTAAAGG	AAATTTTGAA	1920
AATGCTT	TTG	AGATTGCAGG	AGCCAGTATT	CTACTAGAAT	TTATACCAGA	ACTTTTAATA	1980
CCTGTAG	TTG	GAGCCTTTTT	ATTAGAATCA	TATATTGACA	АТАААААТАА	AATTATTAAA	2040
ACAATAG	ATA	ATGCTTTAAC	TAAAAGAAAT	GAAAAATGGA	GTGATATGTA	CGGATTAATA	2100
GTAGCGC	TAA	GGCTCTCAAC	AGTTAATACT	CAATTTTATA	CAATAAAAGA	GGGAATGTAT	2160
AAGGCTI	TAA	ATTATCAAGC	ACAAGCATTG	GAAGAAATAA	TAAAATACAG	ATATAATATA	2220
FATTCTG	AAA	AAGAAAAGTC	AAATATTAAC	ATCGATTTTA	ATGATATAAA	TTCTAAACTT	2280
AATGAGG	GTA	TTAACCAAGC	TATAGATAAT	ATAAATAATT	TTATAAATGG	ATGTTCTGTA	2340
ICATATI	TAA	TGAAAAAAAT	GATTCCATTA	GCTGTAGAAA	AATTACTAGA	CTTTGATAAT	2400
					AATTATATTT		2460
		_			CCATTATGCC		2520
FCAATAT	ATA	CCAATGATAC	AATACTAATA	GAAATGTTTA	ATAAATATAA	TAGC	2574

(2) INFORMATION FOR SEQ ID NO: 28:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2574 base pairs

 - (B) TYPE: nucleic acid (C) STRANDEDNESS: double (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

60	TAATATTATT	TTGATAATAA	AATGATCCTA	TTTTAATTAT	CAATAAATAA	ATGCCAGTTA
120	TAAAATCACA	ATAAAGCTTT	GGGAGATATT	GAGAGGTACG	CTCCATTTGC	ATGATGGAGC
180	GGATTTTAAT	ATAAACCTGA	ACTTTTGGAT	GGAAAGATAT	GGATAATACC	GATCGTATTT
240	TTACTTAAAT	ATGATCCAGA	TGTGAATATT	TAGAGATGTT	GTATTTTAA	AAAAGTTCCG

WO 98/07864 PCT/GB97/02273

- 113 -

ACTAATGATA	AAAAGAATAT	ATTTTACAA	ACAATGATCA	AGTTATTTAA	TAGAATCAAA	300
TCAAAACCAT	TGGGTGAAAA	GTTATTAGAG	ATGATTATAA	ATGGTATACC	TTATCTTGGA	360
GATAGACGTG	TTCCACTCGA	AGAGTTTAAC	ACAAACATTG	CTAGTGTAAC	TGTTAATAAA	420
TTAATCAGTA	ATCCAGGAGA	AGTGGAGCGA	AAAAAAGGTA	TTTTCGCAAA	TTTAATAATA	. 480
TTTGGACCTG	GGCCAGTTTT	AAATGAAAAT	GAGACTATAG	ATATAGGTAT	ACAAAATCAT	540
TTTGCATCAA	GGGAAGGCTT	CGGGGGTATA	ATGCAAATGA	AGTTTTGCCC	AGAATATGTA	600
AGCGTATTTA	ATAATGTTCA	AGAAAACAAA	GGCGCAAGTA	TATTTAATAG	ACGTGGATAT	660
TTTTCAGATC	CAGCCTTGAT	ATTAATGCAT	GAACTCATCC	ACGTCCTCCA	CGGTCTCTAC	720
GGTATCAAAG	TAGACGACCT	CCCGATCGTC	CCGAACGAAA	AAAAATTCTT	CATGCAGAGC	780
ACCGACGCAA	TCCAGGCAGA	AGAACTCTAC	ACCTTCGGTG	GTCAGGACCC	GAGCATCATC	. 840
ACCCCGAGCA	CCGACAAAAG	CATCTACGAC	AAAGTCCTCC	AGAACTTCCG	TGGTATCGTC	900
GACCGTCTCA	ACAAAGTCCT	CGTCTGCATC	AGCGACCCGA	ACATCAACAT	CAACATCTAC	960
AAAAACAAAT	TCAAAGACAA	ATACAAATTC	GTCGAAGACA	GCGAAGGTAA	ATACAGCATC	1020
GACGTCGAGA	GCTTCGACAA	ACTCTACAAA	AGCCTCATGT	TCGGTTTCAC	CGAAACCAAC	1080
ATCGCAGAAA	ACTACAAAAT	CAAAACCCGT	GCAAGCTACT	TCAGCGACAG	CCTCCCGCCG	1140
GTCAAAATCA	AAAACCTCCT	CGACAACGAA	ATCTACACCA	TCGAAGAAGG	TTTCAACATC	1200
AGCGACAAAG	ACATGGAAAA	AGAATACCGT	GGTCAGAACA	AAGCAATCAA	CAAACAAGCT	1260
TACGAAGAAA	TCAGCAAAGA	ACACCTCGCA	GTCTACAAAA	TCCAGATGTG	CAAAAGCGTC	1320
AAAGCACCGG	GTATCTGCAT	CGACGTTGAC	AACGAAGACC	TCTTCTTCAT	CGCAGACAAA	1380
AACAGCTTCA	GCGACGACCT	CAGCAAAAAC	GAACGTATCG	AATACAACAC	CCAGAGCAAC	1440
TACATCGAAA	ACGACTTCCC	GATCAACGAA	CTCATCCTCG	ACACCGACCT	CATCAGCAAA	1500
ATCGAACTCC	CGAGCGAAAA	CACCGAAAGC	CTCACCGACT	TCAACGTTGA	CGTCCCGGTC	1560
TACGAAAAAC	AGCCGGCAAT	CAAAAAAATC	TTCACCGACG	AAAACACCAT	CTTCCAGTAC	1620
CTCTACAGCC	AGACCTTCCC	GCTAGATATA	AGAGATATAA	GTTTAACATC	TTCATTTGAT	1680
GATGCATTAT	TATTTTCTAA	CAAAGTTTAT	TCATTTTTT	CTATGGATTA	TATTAAAACT	1740
GCTAATAAAG	TGGTAGAAGC	AGGATTATTT	GCAGGTTGGG	TGAAACAGAT	AGTAAATGAT	1800
TTTGTAATCG	AAGCTAATAA	AAGCAATACT	ATGGATAAAA	TTGCAGATAT	ATCTCTAATT	1860
GTTCCTTATA	TAGGATTAGC	TTTAAATGTA	GGAAATGAAA	CAGCTAAAGG	AAATTTTGAA	1920
AATGCTTTTG	AGATTGCAGG	AGCCAGTATT	CTACTAGAAT	TTATACCAGA	ACTTTTAATA	1980
CCTGTAGTTG	GAGCCTTTTT	ATTAGAATCA	TATATTGACA	ATAAAAATAA	AATTATTAAA	2040
ACAATAGATA	ATGCTTTAAC	TAAAAGAAAT	GAAAAATGGA	GTGATATGTA	CGGATTAATA	2100
GTAGCGCAAT	GGCTCTCAAC	AGTTAATACT	CAATTTTATA	CAATAAAAGA	GGGAATGTAT	2160
AAGGCTTTAA	-ATTATCAAGC	ACAAGCATTG-	-GAAGAAATAA-	TAAAATACAG	ATATAATATA	2220
TATTCTGAAA	AAGAAAAGTC	AAATATTAAC	ATCGATTTTA	ATGATATAAA	TTCTAAACTT	2280

PCT/GB97/02273

WO 98/07864

- 114 -

AATGAGGGTA	TTAACCAAGC	TATAGATAAT	ATAAATAATT	TTATAAATGG	ATGTTCTGTA	2340
TCATATTTAA	TGAAAAAAAT	GATTCCATTA	GCTGTAGAAA	AATTACTAGA	CTTTGATAAT	2400
ACTCTCAAAA	AAAATTTGTT	ATATTATA	GATGAAAATA	AATTATATTT	GATTGGAAGT	246
GCAGAATATG	AAAAATCAAA	AGTAAATAAA	TACTTGAAAA	CCATTATGCC	GTTTGATCTT	2520
מדמדמתמטד	CCAATGATAC	AATACTAATA	GAAATGTTTA	АТАААТА	TAGC	2574

CLAIMS

- 1. A polypeptide comprising first and second domains, wherein said first domain is adapted to cleave one or more vesicle or plasma-membrane associated proteins essential to exocytosis, and wherein said second domain is adapted (i) to translocate the polypeptide into a cell or (ii) to increase the solubility of the polypeptide compared to the solubility of the first domain on its own or (iii) both to translocate the polypeptide into a cell and to increase the solubility of the polypeptide compared to the solubility of the first domain on its own, said polypeptide being free of clostridial neurotoxin and free of clostridial neurotoxin precursor that can be converted into toxin by proteolytic action.
- 2. A polypeptide according to Claim 1 wherein said first domain comprises a clostridial toxin light chain.
- 3. A polypeptide according to Claim 1 wherein said first domain comprises a fragment or variant of a clostridial toxin light'chain.
- 4. A polypeptide according to Claim 2 or 3 wherein the clostridial toxin is a botulinum toxin.
- 5. A polypeptide according to any preceding claim wherein the first domain exhibits endopeptidase activity specific for a substrate selected from one or more of SNAP-25, synaptobrevin/VAMP and syntaxin.
- 6. A polypeptide according to any preceding claim wherein said second domain comprises a clostridial toxin heavy chain H_N portion.
- 7. A polypeptide according to any of Claims 1-5 wherein said second domain comprises a fragment or variant of a clostridial toxin heavy chain H_N portion.
- 8. A polypeptide according to Claim 6 or 7 wherein the clostridial toxin is a

botulinum toxin.

- 9. A polypeptide according to any of Claims 1-8 further comprising a third domain adapted for binding of the polypeptide to a cell, by binding of the third domain directly to a cell or by binding of the third domain to a ligand or to ligands that bind to a cell.
- 10. A polypeptide according to Claim 9 wherein said third domain is for binding the polypeptide to an immunoglobulin.
- 11. A polypeptide according to Claim 10 wherein said third domain is a tandem repeat synthetic IgG binding domain derived from domain β of Staphylococcal protein A.
- 12. A polypeptide according to Claim 9 wherein said third domain comprises an amino acid sequence that binds to a cell surface receptor.
- 13. A polypeptide according to Claim 12 wherein said third domain is insulin-like growth factor-1 (IGF-1).
- 14. A polypeptide according to any preceding claim comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and a portion designated H_N of a botulinum toxin heavy chain.
- 15. A polypeptide according to Claim 14 wherein one or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type A.
- 16. A polypeptide according to Claim 15 wherein the botulinum toxin type A light chain variant has at residue 2 a glutamate, at residue 26 a lysine and at residue 27 a tyrosine.

- 17. A polypeptide according to Claim 14 wherein one or both of (a) the toxin light chain or fragment or variant of toxin light chain and (b) the portion of the toxin heavy chain are of botulinum toxin type B.
- 18. A polypeptide according to any of Claims 1-13 comprising a botulinum toxin light chain or a fragment or a variant of a botulinum toxin light chain and at least 100 N-terminal amino acids of a botulinum toxin heavy chain.
- 19. A polypeptide according to Claim 18 comprising a botulinum toxin type B light chain, or a fragment or variant thereof, and 107 N-terminal amino acids of a botulinum toxin type B heavy chain.
- 20. A polypeptide according to Claim 15 or 16 comprising at least 423 of the N-terminal amino acids of botulinum toxin type A heavy chain.
- 21. A polypeptide according to Claim 20 comprising a botulinum toxin type A light chain and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
- 22. A polypeptide according to Claim 20 comprising a botulinum toxin type A light chain variant wherein residue 2 is a glutamate, residue 26 is a lysine and residue 27 is a tyrosine, and 423 N-terminal amino acids of a botulinum toxin type A heavy chain.
- 23. A polypeptide according to Claim 17 comprising at least 417 of the N-terminal amino acids of botulinum toxin type B heavy chain.
- 24. A polypeptide according to Claim 23 comprising a botulinum toxin type B light chain and 417 N-terminal amino acids of a botulinum toxin type B heavy chain.
- 25. A polypeptide according to any of Claims 14-24 lacking a portion designated

H_c of a botulinum toxin heavy chain.

- 26. A polypeptide comprising a botulinum toxin light chain and a fragment of a botulinum toxin heavy chain, said fragment being not capable of binding to cell surface receptors.
- 27. A polypeptide according to Claim 26 lacking an intact portion designated $H_{\rm c}$ of a botulinum toxin heavy chain.
- 28. A polypeptide according to any preceding claim comprising a variant of a clostridial toxin and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin.
- 29. A polypeptide according to Claim 28 comprising a variant of a clostridial toxin light chain and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin light chain.
- 30. A polypeptide according to Claim 28 or 29 comprising a variant of a clostridial toxin heavy chain H_N portion and further comprising a site for cleavage by a proteolytic enzyme, which cleavage site is not present in the native toxin heavy chain H_N portion.
- 31. A polypeptide according to Claim 28, 29 or 30 obtainable by modification of a DNA encoding the polypeptide so as to introduce one or more nucleotides coding for the cleavage site.
- 32. A fusion protein comprising a fusion of (a) a polypeptide according to any of Claims 1-31 with (b) a second polypeptide being a polypeptide or oligopeptide adapted for binding to an affinity matrix so as to enable purification of the fusion protein using said matrix.

^{33.} A fusion protein according to Claim 32 wherein said second polypeptide is

adapted to bind to a chromatography column, such as an affinity matrix of glutathione Sepharose.

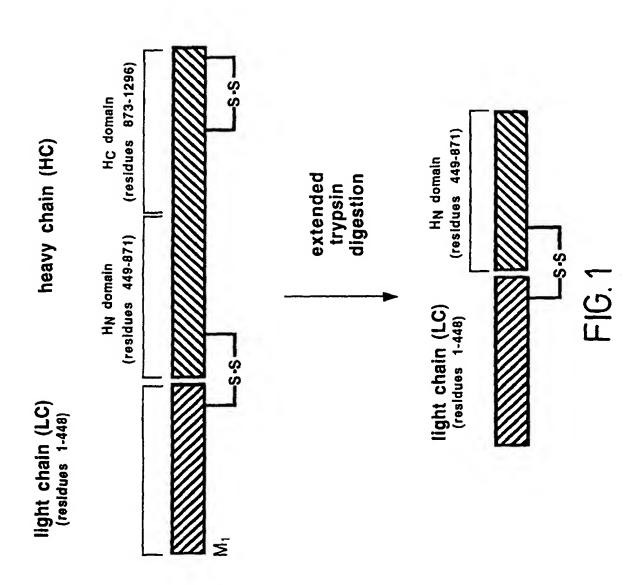
- 34. A fusion protein according to Claim 32 or 33 wherein a specific protease cleavage site is incorporated between the first and second polypeptides, said protease site enabling proteolytic separation of first and second polypeptides.
- 35. A composition comprising a derivative of a clostridial toxin, said derivative retaining at least 10% of the endopeptidase activity of the botulinum toxin, said derivative further being non-toxic *in vivo* due to its inability to bind to cell surface receptors, and wherein the composition is free of any component, such as toxin or a further toxin derivative, that is toxic *in vivo*.
- 36. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for use as a positive control in a toxin assay.
- 37. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for use as a vaccine against clostridial toxin.
- 38. A composition according to Claim 35 or a polypeptide according to any of Claims 1-31 or a fusion protein according to Claim 32, 33 or 34 for *in vivo* use.
- 39. A pharmaceutical composition comprising a composition according to Claim 35, a polypeptide according to any of claims 1-31 or a fusion protein according to Claim 32, 33 or 34, in combination with a pharmaceutically acceptable carrier.
- 40. A nucleic acid encoding a polypeptide or a fusion protein according to any of Claims 1-34.
- 41. A nucleic acid encoding a polypeptide or a fusion protein according to Claim

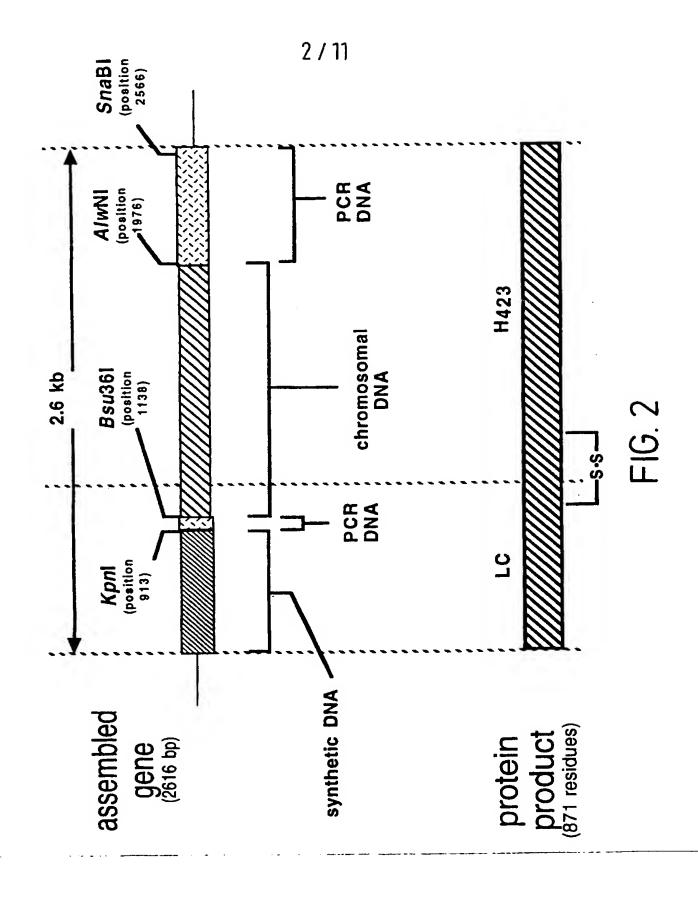
40 and comprising nucleotides encoding residues 1-448 of a botulinum toxin type A light chain.

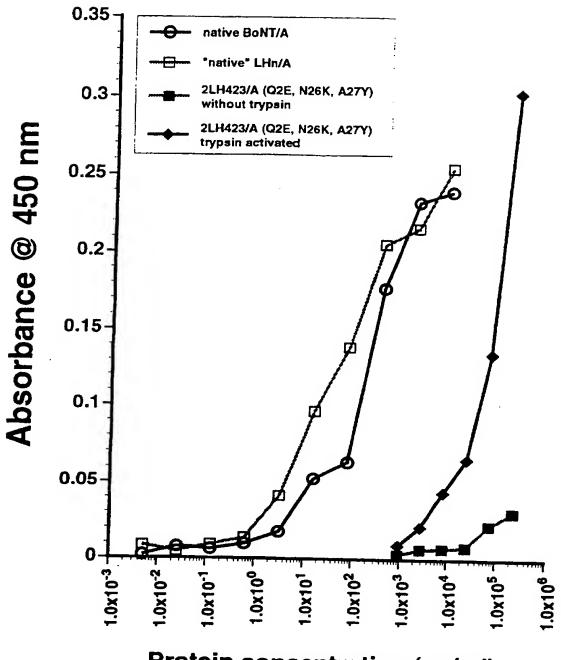
- 42. A nucleic acid according to Claim 40 or 41 comprising nucleotides encoding residues 1-423 of a botulinum toxin type A heavy chain H_N domain.
- 43. A nucleic acid encoding a polypeptide or a fusion protein according to Claim 40 and comprising nucleotides encoding residues 1-470 of a botulinum toxin type B light chain.
- A nucleic acid encoding a polypeptide or a fusion protein according to Claim 44. 40 or 43 comprising nucleotides encoding residues 1-417 of a botulinum toxin type B heavy chain H_N domain.
- 45. A nucleic acid according to any of Claims 40-44 comprising nucleotides ncoding a restriction endonuclease cleavage site not present in native clostridial toxin sequence.
- A nucleotide according to Claim 45 obtainable by modification of a 46. nucleotide encoding a polypeptide or fusion protein according to any of claims 1-34 so as to introduce said cleavage site.
- A DNA according to any of claims 40-46. 47.
- A DNA selected from SEQ ID No:s 1, 8, 10, 12, 14, 16, 18, 23 and 24. 48.
- A method of manufacture of a polypeptide according to any of Claims 1-31 49. comprising expressing in a host cell a nucleic acid according to any of Claims 40-48 and recovering the polypeptide.
- A method of manufacture of a polypeptide according to any of Claims 1-31 50. comprising expressing in a host cell a nucleic acid encoding a fusion protein

according to Claim 32, 33 or 34, purifying the fusion protein by eluting the fusion protein through an affinity matrix adapted to retain the fusion protein and eluting through said matrix a ligand adapted to displace the fusion protein, and recovering the fusion protein.

- 51. A method of manufacture according to Claims 49 or 50 in which the nucleic acid is DNA.
- 52. A cell expressing a polypeptide or fusion protein according to any of Claims 1-34.

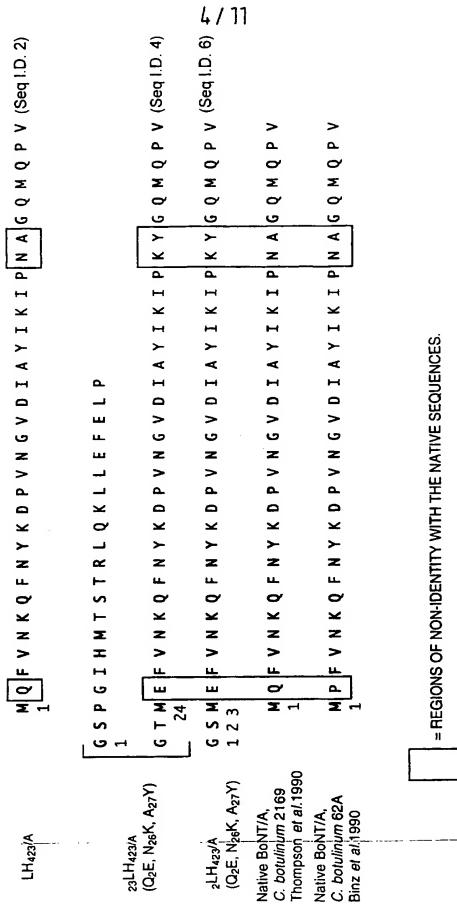






Protein concentration (ng/ml)

FIG. 3



F1G. 4

tgc gat ggg GCA TTA AAT GAT TTA C D G |A L N D L 1321/441

TCA TTA GAT AAA GGA TAC AAT AAG agc gct gat ggg GCA TTA AAT GAT TTA TGT ATC AAA

S L D K G Y N K S A D G A L N D L C I K ggt cgt G R Factor Xa procease mouf FIG. 6 1321/441 TCA TTA GAT AAA GGA TAC AAT AAG atc gaa S L D K G Y N K I E Eco47 III

6/11

CAG ACA
Q T
TAT TGC GAC AGG D R AGG CCT R P 66A 6 CCT ATG M TCT S TGT C 6CG A GAG E AAG K GTG V AGG R CTG L ATT I TTC F CGG R AGG R AGC AGT S S TAT Y CA6 AGG R CTA 2617/873 TTT ACT F T 2677/893 GAT GCT D A 2737/913 GGC TCC G S 2797/933 TGT GAT C D 2857/953 GCT TAG A stop . GAA TCT ACA S T CTG GTG L V GGG TAT G Y CGG AGC GCT A TCA S TTA L GAG E ACA T TTC F TTA L GCT A AAG K 000 000 000 000 000 9 8 8 AGA R AAG K 16C C 999 AAC N GAG E CCT AAG F GAT D AAT N CTC L ר כל GAT D ACG T 6TG V TAT Y 2587/863 TAC GTA (Y V | 1 2647/883 CCG GAG (P E | 2707/903 GGC TTT (G F | 2767/923 GGT ATC (G I 2 2827/943 GCA CCC A

F1G. 7

AAT N AGG R A F ACA T TCT S AAG K GAT D ATT ATT I GAT D TAT Y GAA E TCT S 2617/873 TTT ACT F T 2677/893 TAT CAA TCT ACA S T TCA GGC S G TT A TT T CAA ATA Q I TTA L AGA R CAA AGA TTA L AAA K AAT N BA E GAT D FT > GAT 2587/863 TAC GTA Y Y V 2647/883 TCT AAA H S K 2707/903 ATT AAG

F1G. 8

	AAT	z	:	AAA	×	:	CGA	œ		GAA	ىيا	l	CAA	· C	v	رين	, V		AAA	×	:		
	CTG		ı	AAC	z	:	CAA	0	,	SCA	<	:	GAA	u.	ı	AAC	z	:	F)	٥	:		
	9	ی	,	711	<u>.</u>		GAA	ш		TTA	_		AAA	~	:	CGA	~	:	GAA	u.	ı		
																		y					
	AAG	<u></u>	:	AAC	z		AAC	z		AAC	z		110	L.		GAA		,	TTA	_	ı		
	AT		ı	SAC	۵		TTA	ب		FJ.	⋖		AAA	×		GAA	w	ı	E	_	ī		
	GAA	w		S S	4		CCT	Φ.		CAA	0	,	SAC	۵		TA			SCT	⋖	m		
7/873	ACT	—	7/893	GAA	w	7/913	TA	ب	7/933	AGC	S	7/953	GTA	>	7,973	AAC	z	7993	AGC	S	/101	TAG	*
2617	T	u.	267	GAT	۵	273	CAT	×	279	S	ط	2857	AAA	¥	2917	L)	۵	2977	CAA	0	3037	GAC	۵
																						GTA	
	TCT	S		CAA	0	,	ATC	H		GAT	0		S	4		CAT	I		S	۵.		AAA	¥
	TA	۔		939	4		GAG	ш		AAA	¥		SAG	0	,	TA	_		GAC			923	۵
	TA	_		TAT	>	ı	TAT	> -		TA	_		₽ U	⋖		ATC	Н		GAT	0		939	∢
	AGA	œ		CAT	I		H	u.		AGT	S		GAT	0		GAG	u		AAA	×		CAG	0
	B	0	•	SCT CT	4		939	4		S	0	,	AAT	z		TAT	>-		TA	_		SCT	∢
	AAT	z		ScA	¥		AAC	z		ATC	H		CTA	_		E	ட		AGT	S		GAT	۵
_	GAT	۵		GGT	ی	_	SA	0		E	<u>u</u>		AAG	×		S	⋖		SA	0	m	AAT	z
7/86	STA	>	7/88	S	م	7/90	₹	0	7/923	ပ္ပ	⋖	7/943	AAA	¥	1/963	AAC	z	//983	ATC	H	/100	CTA	_
2587	TAC	> -	264.	7	S	2707	SAA	ш	2767	AAC	z	2827	PS L	4	2887	B	0	2947	Ĕ	L	3007	AAG CTA AAT GAT GCT CAG GC	×

F16.0

8/11 LH₄₂₃/A (R1)(R2)(R3)GSTلـى۔ synthetic IgG binding domain R1 cleavage (R2)(R3)cleavage and activation by R2 anti-receptor/marker antibody Intracellular reduction LC H_{423} SH SH

FIG. 10

$LH_{423}/A^{9/11}$

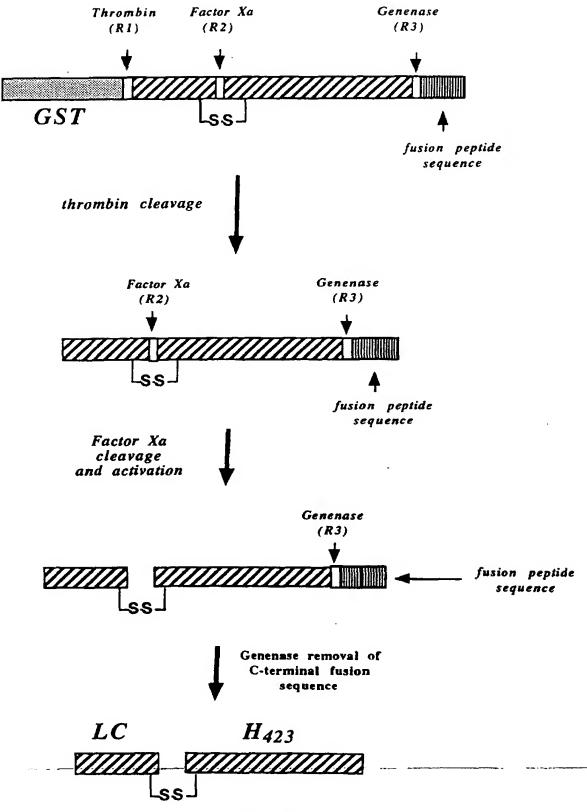


FIG. 11

10/11

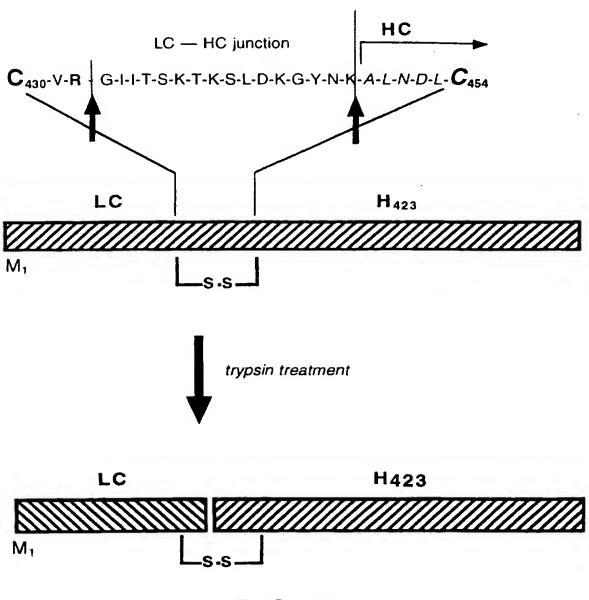
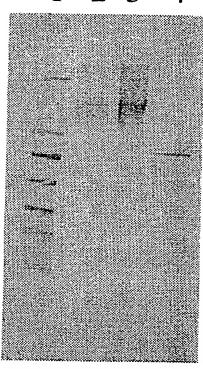


FIG. 12

11 / 11

Panel A. 1 2 3 4



Panel B. 1 2 3 4

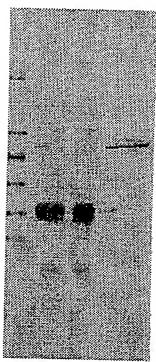


FIG. 13

TIONAL SEARCH REPORT

Inten unal Application No PCT/GB 97/02273

A. CLASSIFICATION OF SUBJECT MATTER IPC 6 C12N15/31 C12N1/21 A61K39/08

C12P21/02

CO7K14/33

A61K38/16

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols) IPC 6 C12N C12P A61K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

Category *	Citation of document, with indication, where appropriate, of the relevant passages	
		Relevant to claim No.
X	WO 96 12802 A (OPHIDIAN PHARM INC ;WILLIAMS JAMES A (US); PADHYE NISHA V (US); KI) 2 May 1996 see the whole document	1-52
X	KURAZONO H ET AL: "Minimal essential *domains* specifying toxicity of the *light* *chains* of tetanus toxin and botulinum neurotoxin type A." J BIOL CHEM, JUL 25 1992, 267 (21) P14721-9, UNITED STATES, XP002047910 see table II	1-52

X Further documents are listed in the continuation of box C.	Patent family members are listed in annex.
"Special categories of cited documents: "A" document defining the general state of the art which is not considered to be of particular relevance. "E" earlier document but published on or after the international filling date. "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified). "O" document referring to an oral disclosure, use, exhibition or other means. "P" document published prior to the international filling date but later than the priority date claimed.	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "&" document member of the same patent family
Date of the actual completion of the international search	Date of mailing of the international search report
9 December 1997	3 0. 01. 98
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 MV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo ni, Fax: (+31-70) 340-3016	Authorized officer Hillenbrand, G

Interaction No PCT/GB 97/02273

		PCT/GB 97	//022/3
C.(Continu	ation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages		Relevant to claim No.
X	LI ET AL: "A SINGLE MUTATION IN THE RECOMBINANT LIGHT CHAIN OF TETANUS TOXIN ABOLISHES ITS PROTEOLYTIC ACTIVITY AND REMOVES THE TOXICITY SEEN AFTER RECONSTITUTION WITH NATIVE HEAVY CHAIN" BIOCHEMISTRY, vol. 33, no. 22, 1994, pages 7014-7020, XP002015938 see the whole document		1
A	BINZ T ET AL: "THE COMPLETE SEQUENCE OF BOTULINUM NEUROTOXIN TYPE A AND COMPARISON WITH OTHER CLOSTRIDIAL NEUROTOXINS" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 265, no. 16, 5 June 1990, pages 9153-9158, XP002009348 see the whole document	·	1,26,35
	·		
٠			

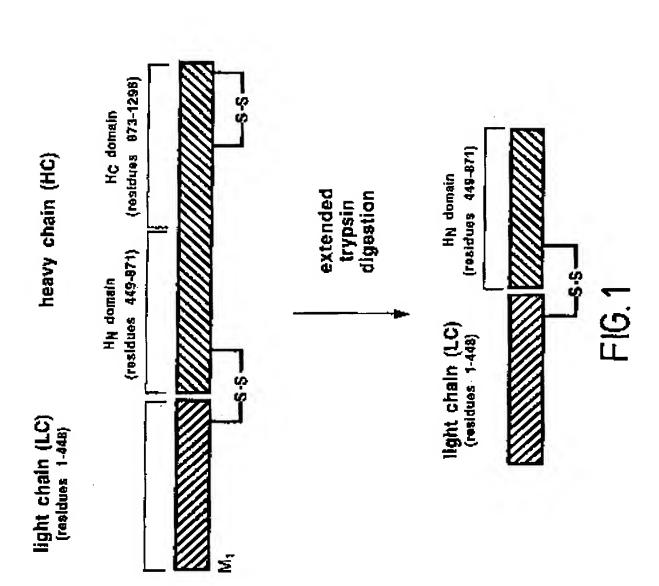
INTERATIONAL SEARCH REPORT

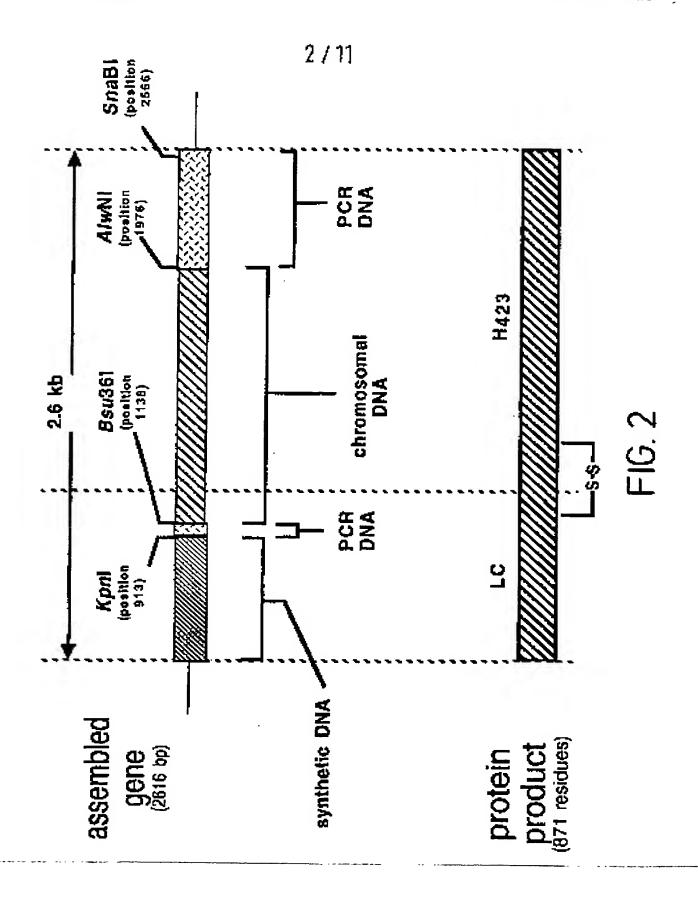
Intel. .onal Application No PCT/GB 97/02273

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9612802 A	02-05-96	US 5601823		11-02-97
		US 5196193	Α	23-03-93
		AU 3968395	Α	15 - 05-96
		EP 0796326	Α	24-09-97
		FI 971732	Α	23-06-97
		NO 971868	Α	24-06-97
		PL 320214	Α	15-09-97
		ZA 9508990	Α	15-05-96
		AU 6653894	Α	04-07-94
		CA 2150935	Α	23-06-94
		EP 0671902	Α	20-09-95
		WO 9413264	Α	23-06-94
		US 5466672	Α	14-11-95
		US 5599539	Α	04-02-97
		AU 638786	В	08-07-93
		AU 6895191		31-05-91
		EP 0498854	Α	19 - 08-92
	•	WO 9106306	Α	16-05-91
		US 5443976	Α	22-08-95
		US 5340923	Α	23-08-94

		•	
€			

_





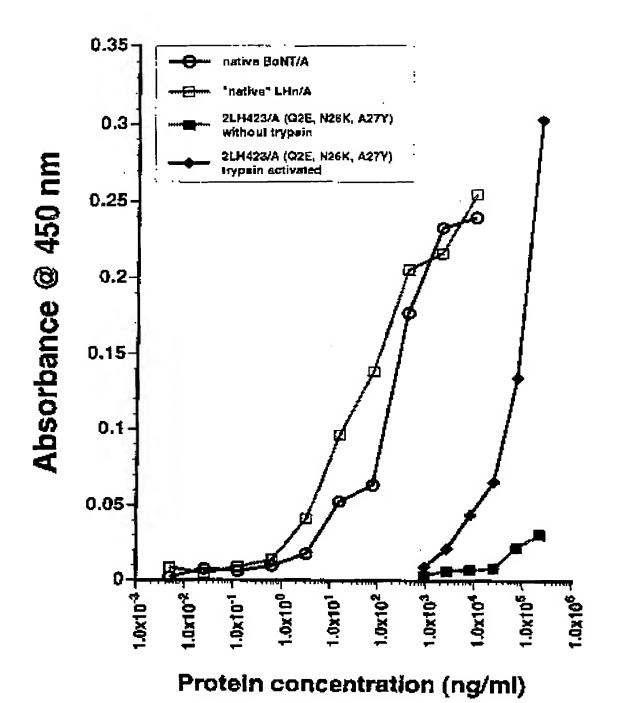


FIG. 3

£ī			4/1	7			
Q P V (Seq I.D. 2)		(Seq 1.D. 4)	(Seq I.D. 6)		٠		
(Seq		(Sec	(Sed				
>		V 9 9		>	>		
<u> </u>		Δ.	> ⊶	₫.	<u> </u>		
ď			Ŏ	0	ď		
芝		3 :	Ç.\ X	⊙ .	∑ Z		
C		0 9	_G_	<u> </u>			
IKIPNAG		<u> </u>	_	₹	₩ G		
2		> ×	×	Z	Z		
		-	Δ,	<u> </u>	2		
H		×	×	H	¥		
×		×		×	¥		
<u></u> -		H	⊢	H	Η,		တ်
*		≻	≯- -⊀	>- ≪Z	≻ -		S
ч	Δ.	H	Ħ	H	×		Ñ
H	_	⇔	_	Ð	۵		렰
>	144	>	>	>	>		SE
ی	14.	Ġ	Φ	G	9		₹
7	ш	Z	2	2	Z		₽¥
٥ ٥	اس ت	>	>	>	>		낕
<u>-</u>	-	9	م	<u>6</u>	<u>.</u>		E
7	<u>о</u> , Ж	*	*	₹	¥		臣
F X Y	_	≻ -	>-	>	>-		3
Z	∞ :	Z	Z	Z	Z		E
ш	!	4-	£L_	LL .	QFNY		돏
~	41	0		Ċ	0		₽
2	⊢	≥ ∠	×	×	~		₹
~	=	-	=	=	~		Z
<u> </u>	Ä	u.	LL.	11.	11.		O to
MQF VNK	- G S P G I H H T 1	GTHEFVNK - 24	G S M E F V N K	1 0 F V N K	M Y X X X X X X X X X X X X X X X X X X		= REGIONS OF NON-IDENTITY WITH THE NATIVE SEQUENCES.
ΣH	•	24 x 4	Ξm	≖ ल	₹	4	<u> </u>
•	V3	⊢ ''	G S M				配
							· · · · ·
				ð			
		Σ	بخ	85	-eC*		
		4,	₹,	₹ <u>2</u>	ર્ છે ટ	2	
₹		~ 	8 X	- Z-5-c-	<u> </u>		
LH423/A	H189	(O ₂ E. N ₂₆ K, A ₂₇ Y)	2LH423/A (QzE, NzgK, AzzY)	Native BoNT/A, C. botulinum 2169 Thompson et al.1990	Native BoNT/A, C. <i>botulinum</i> 62A Binz et al 1890	-	
7	<u> </u>	Š	Ŧ3	trive box	Five Sol	. !	
	r	-	,, -	₽u≓	光 ら声	i	

7 913

1321/441 TCA TTA GAT ANA GGA TAC AAT AAG ogc gct got ggg GCA TTA AAT GAT TTA TGT ATC AAA S L D K G Y N K S A D G A L N D L C I K H-chain F G S	1351/451 TCA TTA GAT AAA GGA TAC AAT AAG OTC goo ggt cgt tgc gat ggg GCA TTA AAT GAT TTA S L D K G Y N K I E G R C D G A L N D L Factor Xa Protease molif FIG. 6
--	--

6/11

GAC AGG D R CAG ACA
Q T
TAT TGC AGG R TGT GGA C G GAG ATG E M CCT P rct s 900 A AAG K GTG Y AGG R CTG AGG AGG R TTC F CGG R 5 5 5 6 AĞT S 2617/873
FTT ACT GAA
F T E
2677/893
GAT GCT CTT (
0 A L
2737/913
GGC TCC AGC (
6 S S
2797/933
TGT GAT CTA (
C D L 6
2857/953 GCT GAA GTG Y র ই ⊏ AGC S TÅŢ Y 7CT 5 1 CGG ନ୍ଦିର ଜ Ħ Ħ Ħ TC F GÅG E **₽**C₽ 76C C AAG K Ω A Δ ង្គ AGA R 550 AX 750 9 A 6.4G E ٦<u>6</u> د AAC N GAT D **≸** × בָּל ה GAT D ACG T 616 V TAT Y CTC L

=16.7

CAT H AGG ACA GAT A AG ATT J ATT 1 GAT D TAT Y a a a TCT 2617/873 TTT ACT F T 2677/893 TAT CAA Y Q TCT ACA TCA GGC S G ¥ ∐ ... ATA I Ĕ, AGA R 30 ¥ l 30 AGA R 호 장 노 AAT N Ę, **5**4 SAT 0 2587/863 TAC GTA Y V 2647/883 TCT AAA S K

F16.8

•

	AAT	z	444	~		8	∝		GAA	ш		ርልል	o		၁၂	4		AAA	¥				
	C16	_	AAC	z		₹	0		5	-⊄		GAA	щ		AAC	z		Ų	⋖				
	S	u	Ή	ц,		GAA	L L.f		ŢŢ	_		ልልል	×		3	∝		\$€	ш				
		ις.																6C4	⋖				
		¥																TTA	_				
		ы																					
	፲ልፐ	> -	GTA	>		AAC	z		AGC	S		AAC	z		AAC	Z		AAC	z				
	3	F T E 1 2677/893	S	⋖		5	<u>م</u>		\$	<u>~</u>		S	0		TTA	ب		ST.	⋖	در			
£28/	ACT	T 7∕893	SAA	Ų.	7/913	Ë	_	7/933	AGC	₽ 3	7/953	GTA	>	7/973	Ą	2	7/993	AGC	νı	7/101	TAG	•	_
2617	E	F 2673	GAT	_	273	F	Ŧ	523	ij	_	285	444	¥	291	C,	<u>~</u>	297	CAA	O	383	9	Ö	6.9
	A A	-	CAC.	Ŧ		Ĭ			SAC	-		<u>5</u>	۵.		TTA	ے		AGC	v		Ğ	>	2:
	77.	- S	CAA	0		ATC	Н		B	<u> </u>		976	⋖		CAT	I		Ş	⊅ -		AAA	¥	ш.
		_																			y		
	Ĭ	_	TAT	>-		TAT	>-		TT.	_		£3	-<		ATC	H		GAT	<u></u>		9	_	
	404	€	3	Ŧ		H	ш.		AGT	Ś		GAT	-		GAG	LJ.		AAA	ᆇ			0	
	ð	Ó.																			SCT	_	
	AAT	z	Ş	4		AAC	×		ĄŢ	H		Ë			1	Ŀ		AGT	S		GAT	0	
	GAT	•	<u>199</u>	ي ا		3	0		Ë	4		AAG	×		9	₹		3	¢	Ē	AAT	22	
7/863	GTA	Y V D N		_	SE SE	GAA	0	7,923	y	- 4	7.943	AAA	~	7/963	¥	z	7/983	ATC	H	7/106	CIA		
2587	TAC	7. K.	1	S	2797	GAA	بعر	276	AAC	æ	282	ָּבֶּ	-4	288	3	0	7	Ĕ	ഥ	366	3	¥	
																							_

8 / 11 LH₄₂₃/A

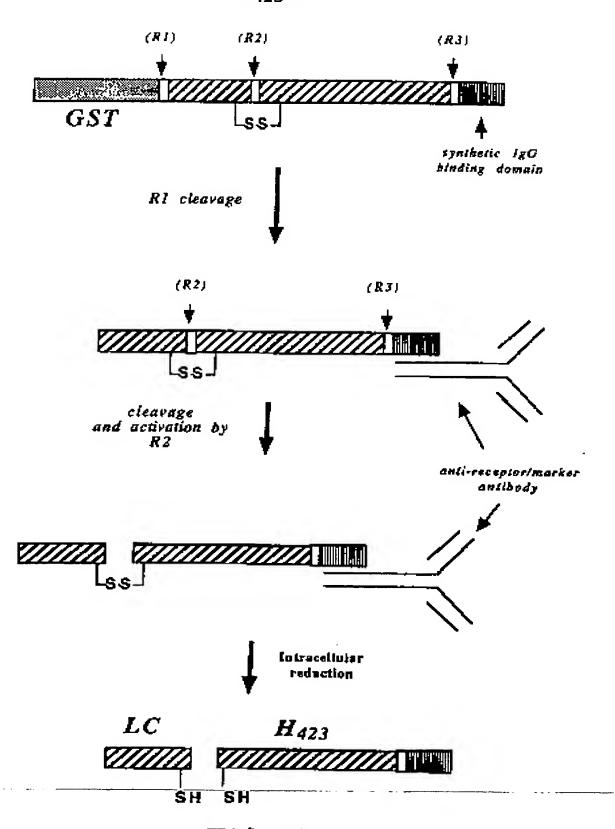
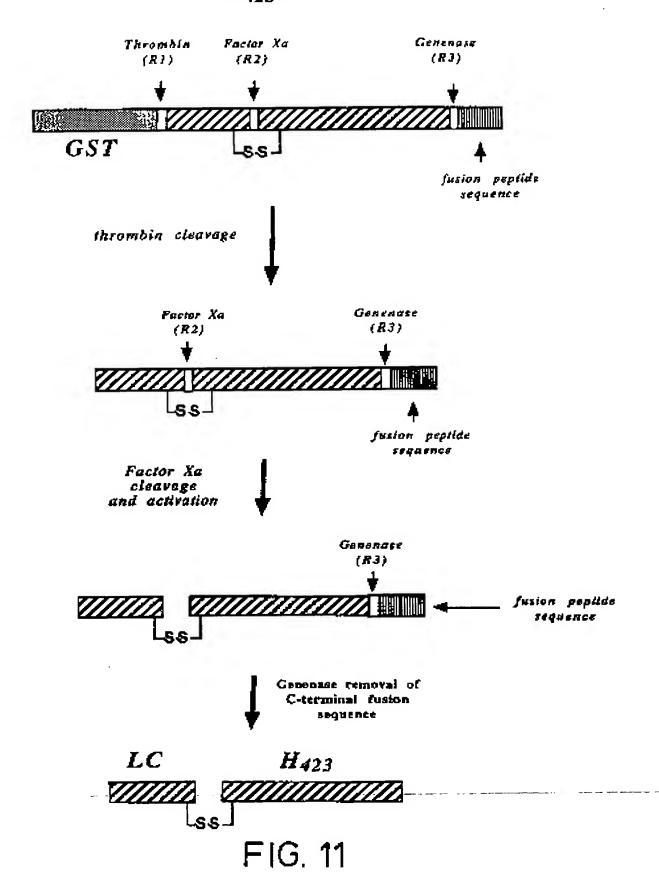


FIG. 10

$LH_{423}/A^{9/11}$



10/11

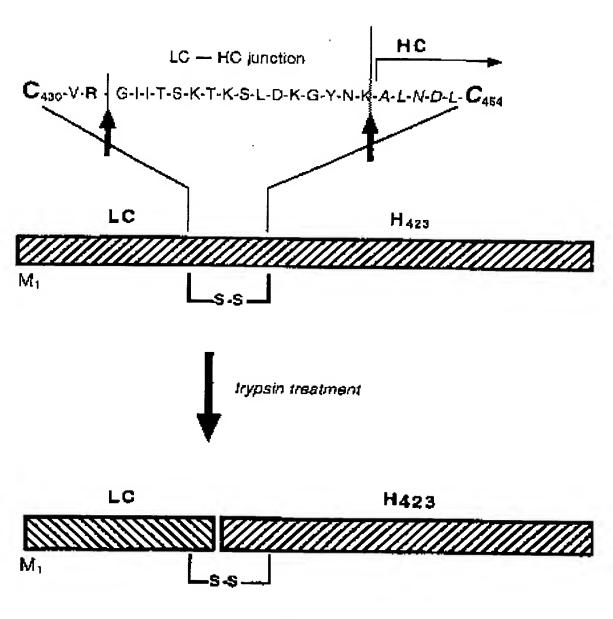
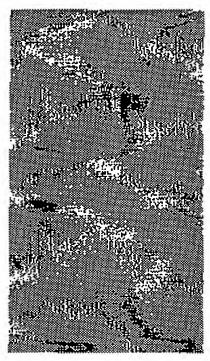


FIG. 12

Panel A. 1 2 3 4



Panel B. 1 2 3 4



FIG. 13

			and the second	
		V		
•				